



HAL
open science

Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution

Hedy Attouch, Szilárd Csaba László

► **To cite this version:**

Hedy Attouch, Szilárd Csaba László. Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution. 2021. hal-03207541

HAL Id: hal-03207541

<https://hal.science/hal-03207541>

Preprint submitted on 25 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution

Hedy Attouch · Szilárd Csaba László

the date of receipt and acceptance should be inserted later

Abstract In a Hilbertian framework, for the minimization of a general convex differentiable function f , we introduce new inertial dynamics and algorithms that generate trajectories and iterates that converge fastly towards the minimizer of f with minimum norm. Our study is based on the non-autonomous version of the Polyak heavy ball method, which, at time t , is associated with the strongly convex function obtained by adding to f a Tikhonov regularization term with vanishing coefficient $\epsilon(t)$. In this dynamic, the damping coefficient is proportional to the square root of the Tikhonov regularization parameter $\epsilon(t)$. By adjusting the speed of convergence of $\epsilon(t)$ towards zero, we will obtain both rapid convergence towards the infimal value of f , and the strong convergence of the trajectories towards the element of minimum norm of the set of minimizers of f . In particular, we obtain an improved version of the dynamic of Su-Boyd-Candès for the accelerated gradient method of Nesterov. This study naturally leads to corresponding first-order algorithms obtained by temporal discretization. In the case of a proper lower semicontinuous and convex function f , we study the proximal algorithms in detail, and show that they benefit from similar properties.

Keywords Accelerated gradient methods; convex optimization; damped inertial dynamics; minimum norm solution; Nesterov accelerated gradient method; Tikhonov approximation.

AMS subject classification 37N40, 46N10, 49M30, 65B99, 65K05, 65K10, 90B50, 90C25.

1 Introduction

Throughout the paper, \mathcal{H} is a real Hilbert space which is endowed with the scalar product $\langle \cdot, \cdot \rangle$, with $\|x\|^2 = \langle x, x \rangle$ for $x \in \mathcal{H}$. We consider the convex minimization problem

$$\min \{f(x) : x \in \mathcal{H}\}, \quad (1)$$

where $f : \mathcal{H} \rightarrow \mathbb{R}$ is a convex continuously differentiable function whose solution set $S = \operatorname{argmin} f$ is nonempty. We aim at finding by rapid methods the element of minimum norm of S . As an original aspect of our approach, we start from the Polyak heavy ball with friction dynamic for strongly

H. Attouch
IMAG, Univ. Montpellier, CNRS, Montpellier, France
E-mail: hedy.attouch@umontpellier.fr

S. László
Technical University of Cluj-Napoca, Department of Mathematics, Memorandumului 28, Cluj-Napoca, Romania
E-mail: szilard.laszlo@math.utcluj.ro

convex functions, and then adapt it to treat the case of general convex functions. Recall that a function $f : \mathcal{H} \rightarrow \mathbb{R}$ is said to be μ -strongly convex for some $\mu > 0$ if $f - \frac{\mu}{2} \|\cdot\|^2$ is convex. In this setting, we have the exponential convergence result:

Theorem 1 *Suppose that $f : \mathcal{H} \rightarrow \mathbb{R}$ is a function of class \mathcal{C}^1 which is μ -strongly convex for some $\mu > 0$. Let $x(\cdot) : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of*

$$\ddot{x}(t) + 2\sqrt{\mu}\dot{x}(t) + \nabla f(x(t)) = 0. \quad (2)$$

Then, the following property holds: $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(e^{-\sqrt{\mu}t}\right)$ as $t \rightarrow +\infty$.

Let us see how to take advantage of this fast convergence result, and how to adapt it to the case of a general convex differentiable function $f : \mathcal{H} \rightarrow \mathbb{R}$. The main idea is linked to Tikhonov's method of regularization. It consists in considering the corresponding non-autonomous dynamic which at time t is governed by the gradient of the strongly convex function $f_t : \mathcal{H} \rightarrow \mathbb{R}$

$$f_t(x) := f(x) + \frac{\epsilon(t)}{2} \|x\|^2.$$

Then replacing f by f_t in (2), and noticing that f_t is $\epsilon(t)$ -strongly convex, we obtain the dynamic

$$\text{(TRIGS)} \quad \ddot{x}(t) + \delta\sqrt{\epsilon(t)}\dot{x}(t) + \nabla f(x(t)) + \epsilon(t)x(t) = 0,$$

with $\delta = 2$. (TRIGS) stands shortly for Tikhonov regularization of inertial gradient systems. In order not to asymptotically modify the equilibria, we suppose that $\epsilon(t) \rightarrow 0$ as $t \rightarrow +\infty$. This condition implies that (TRIGS) falls within the framework of the inertial gradient systems with asymptotically vanishing damping. The importance of this class of inertial dynamics has been highlighted by several recent studies [3], [5], [8], [10], [18], [28], [38], which make the link with the accelerated gradient method of Nesterov [35,36].

1.1 Historical facts and related results

In relation to optimization algorithms, a rich literature has been devoted to the coupling of dynamic gradient systems with Tikhonov regularization.

1.1.1 First-order gradient dynamics

For first-order gradient systems and subdifferential inclusions, the asymptotic hierarchical minimization property which results from the introduction of a vanishing viscosity term in the dynamic (in our context the Tikhonov approximation [39,40]) has been highlighted in a series of papers [2], [4], [12], [14], [20], [30], [33]. In parallel way, there is a vast literature on convex descent algorithms involving Tikhonov and more general penalty, regularization terms. The historical evolution can be traced back to Fiacco and McCormick [31], and the interpretation of interior point methods with the help of a vanishing logarithmic barrier. Some more specific references for the coupling of Prox and Tikhonov can be found in Cominetti [29]. The time discretization of the first-order gradient systems and subdifferential inclusions involving multiscale (in time) features provides a natural link between the continuous and discrete dynamics. The resulting algorithms combine proximal based methods (for example forward-backward algorithms), with the viscosity of penalization methods, see [15], [16], [22], [25,26], [33].

1.1.2 Second order gradient dynamics

First studies concerning the coupling of damped inertial dynamics with Tikhonov approximation concerned the heavy ball with friction system of Polyak [37], where the damping coefficient $\gamma > 0$ is fixed. In [13] Attouch-Czarnecki considered the system

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) + \epsilon(t)x(t) = 0. \quad (3)$$

In the slow parametrization case $\int_0^{+\infty} \epsilon(t)dt = +\infty$, they proved that any solution $x(\cdot)$ of (3) converges strongly to the minimum norm element of $\operatorname{argmin} f$, see also [34]. A parallel study has been developed for PDE's, see [1] for damped hyperbolic equations with non-isolated equilibria, and [2] for semilinear PDE's. The system (3) is a special case of the general dynamic model

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) + \epsilon(t)\nabla g(x(t)) = 0 \quad (4)$$

which involves two functions f and g intervening with different time scale. When $\epsilon(\cdot)$ tends to zero moderately slowly, it was shown in [17] that the trajectories of (4) converge asymptotically to equilibria that are solutions of the following hierarchical problem: they minimize the function g on the set of minimizers of f . When $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ is a product space, defining for $x = (x_1, x_2)$, $f(x_1, x_2) := f_1(x_1) + f_2(x_2)$ and $g(x_1, x_2) := \|A_1x_1 - A_2x_2\|^2$, where the A_i , $i \in \{1, 2\}$ are linear operators, (4) provides (weakly) coupled inertial systems. The continuous and discrete-time versions of these systems have a natural connection to the best response dynamics for potential games [14], domain decomposition for PDE's [7], optimal transport [6], coupled wave equations [32].

In the quest for a faster convergence, the following system

$$(\text{AVD})_{\alpha, \epsilon} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) + \epsilon(t)x(t) = 0, \quad (5)$$

has been studied by Attouch-Chbani-Riahi [11]. It is a Tikhonov regularization of the dynamic

$$(\text{AVD})_{\alpha} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) = 0, \quad (6)$$

which was introduced by Su, Boyd and Candès in [38]. When $\alpha = 3$, $(\text{AVD})_{\alpha}$ can be viewed as a continuous version of the accelerated gradient method of Nesterov. It has been the subject of many recent studies which have given an in-depth understanding of the Nesterov acceleration method, see [3], [8], [10], [38]. The results obtained in [11] concerning (5) will serve as a basis for comparison.

1.2 Model results

To illustrate our results, let us consider the case $\epsilon(t) = \frac{c}{t^r}$ where r is positive parameter satisfying $0 < r \leq 2$. The case $r = 2$ is of particular interest, it is related to the continuous version of the accelerated gradient method of Nesterov, with optimal convergence rate for general convex differentiable function f .

1.2.1 Case $r = 2$

Let us consider the (TRIGS) dynamic

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) + \frac{c}{t^2} x(t) = 0, \quad (7)$$

where the parameter $\alpha \geq 3$ plays a crucial role. As a consequence of Theorems 8 and 9 we have

Theorem 2 Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution of (7). We then have the following results:

- i) If $\alpha = 3$, then $f(x(t)) - \min_{\mathcal{H}} f = O\left(\frac{\ln t}{t^2}\right)$ as $t \rightarrow +\infty$.
- ii) If $\alpha > 3$, then $f(x(t)) - \min_{\mathcal{H}} f = O\left(\frac{1}{t^2}\right)$ as $t \rightarrow +\infty$. Further, the trajectory x is bounded, $\|\dot{x}(t)\| = O\left(\frac{1}{t}\right)$ as $t \rightarrow +\infty$, and there is strong convergence to the minimum norm solution:

$$\liminf_{t \rightarrow +\infty} \|x(t) - x^*\| = 0.$$

1.2.2 Case $r < 2$

As a consequence of Theorems 7 and 11, we have:

Theorem 3 Take $\epsilon(t) = 1/t^r$, $\frac{2}{3} < r < 2$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a global solution trajectory of

$$\ddot{x}(t) + \frac{\delta}{t^{\frac{r}{2}}} \dot{x}(t) + \nabla f(x(t)) + \frac{1}{t^r} x(t) = 0.$$

Then, we have fast convergence the values, and strong convergence to the minimum norm solution:

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{\frac{3r}{2}-1}}\right) \text{ and } \liminf_{t \rightarrow +\infty} \|x(t) - x^*\| = 0.$$

These results are completed by showing that, if there exists $T \geq t_0$, such that the trajectory $\{x(t) : t \geq T\}$ stays either in the open ball $B(0, \|x^*\|)$ or in its complement, then $x(t)$ converges strongly to x^* as $t \rightarrow +\infty$. Corresponding results for the associated proximal algorithms, obtained by temporal discretization, are obtained in Section 5.

A remarkable property of the above results is that the rate of convergence of values is comparable to the Nesterov accelerated gradient method. In addition, we have a strong convergence property to the minimum norm solution, with comparable numerical complexity. These results represent an important advance compared to previous works by producing new dynamics for which we have both rapid convergence of values and strong convergence towards the solution of minimum norm. Let us stress the fact that in our approach the fast convergence of the values and the strong convergence towards the solution of minimum norm are obtained for the same dynamic, whereas in the previous works [11], [13], they are obtained for different dynamics obtained for different settings of the parameters. It is clear that the results extend naturally to obtaining strong convergence towards the solution closest to a desired state x_d . It suffices to replace in Tikhonov's approximation $\|x\|^2$ by $\|x - x_d\|^2$. This is important for inverse problems.

1.3 Contents

In section 2, we show existence and uniqueness of a global solution for the Cauchy problem associated with (TRIGS). Then, based on Lyapunov analysis, we obtain convergence rates of the values which are valid for a general $\epsilon(\cdot)$. Section 3 is devoted to an in-depth analysis in the critical case $\epsilon(t) = c/t^2$. Section 4 is devoted to the study of the strong convergence property of the trajectories towards the minimum norm solution, in the case of a general $\epsilon(\cdot)$. Then in Section 5 we obtain similar results for the associated proximal algorithms, obtained by temporal discretization.

2 Convergence analysis for general $\epsilon(t)$

We are going to analyze via Lyapunov analysis the convergence properties as $t \rightarrow +\infty$ of the solution trajectories of the inertial dynamic (TRIGS) that we recall below

$$\ddot{x}(t) + \delta\sqrt{\epsilon(t)}\dot{x}(t) + \nabla f(x(t)) + \epsilon(t)x(t) = 0. \quad (8)$$

Throughout the paper, we assume that t_0 is the origin of time, δ is a positive parameter, and

(H₁) $f : \mathcal{H} \rightarrow \mathbb{R}$ is convex and differentiable, ∇f is Lipschitz continuous on bounded sets.

(H₂) $S := \operatorname{argmin} f \neq \emptyset$. We denote by x^* the element of minimum norm of S .

(H₃) $\epsilon : [t_0, +\infty[\rightarrow \mathbb{R}^+$ is a nonincreasing function, of class \mathcal{C}^1 , such that $\lim_{t \rightarrow \infty} \epsilon(t) = 0$.

2.1 Existence and uniqueness for the Cauchy problem

Let us first show that the Cauchy problem for (TRIGS) is well posed.

Theorem 4 *Given $(x_0, v_0) \in \mathcal{H} \times \mathcal{H}$, there exists a unique global classical solution $x : [t_0, +\infty[\rightarrow \mathcal{H}$ of the Cauchy problem*

$$\begin{cases} \ddot{x}(t) + \delta\sqrt{\epsilon(t)}\dot{x}(t) + \nabla f(x(t)) + \epsilon(t)x(t) = 0 \\ x(t_0) = x_0, \dot{x}(t_0) = v_0. \end{cases} \quad (9)$$

Proof The proof relies on the combination of the Cauchy-Lipschitz theorem with energy estimates. First consider the Hamiltonian formulation of (9) as the first order system

$$\begin{cases} \dot{x}(t) - y(t) = 0 \\ \dot{y}(t) + \delta\sqrt{\epsilon(t)}y(t) + \nabla f(x(t)) + \epsilon(t)x(t) = 0 \\ x(t_0) = x_0, y(t_0) = v_0. \end{cases} \quad (10)$$

According to the hypothesis (H₁), (H₂), (H₃), and by applying the Cauchy-Lipschitz theorem in the locally Lipschitz case, we obtain the existence and uniqueness of a local solution. Then, in order to pass from a local solution to a global solution, we rely on the energy estimate obtained by taking the scalar product of (TRIGS) with $\dot{x}(t)$. It gives

$$\frac{d}{dt} \left(\frac{1}{2} \|\dot{x}(t)\|^2 + f(x(t)) + \frac{1}{2} \epsilon(t) \|x(t)\|^2 \right) + \delta\sqrt{\epsilon(t)} \|\dot{x}(t)\|^2 - \frac{1}{2} \dot{\epsilon}(t) \|x(t)\|^2 = 0.$$

From (H₃), $\epsilon(\cdot)$ is non-increasing. Therefore, the energy function $t \mapsto W(t)$ is decreasing where

$$W(t) := \frac{1}{2} \|\dot{x}(t)\|^2 + f(x(t)) + \frac{1}{2} \epsilon(t) \|x(t)\|^2.$$

The end of the proof follows a standard argument. Take a maximal solution defined on an interval $[t_0, T[$. If T is infinite, the proof is over. Otherwise, if T is finite, according to the above energy estimate, we have that $\|\dot{x}(t)\|$ remains bounded, just like $\|x(t)\|$ and $\|\ddot{x}(t)\|$ (use (TRIGS)). Therefore, the limit of $x(t)$ and $\dot{x}(t)$ exists when $t \rightarrow T$. Applying the local existence result at T with the initial conditions thus obtained gives a contradiction to the maximality of the solution.

2.2 General case

The control of the decay of $\epsilon(t)$ to zero as $t \rightarrow +\infty$ will play a key role in the Lyapunov analysis of (TRIGS). Precisely, we will use the following condition.

Definition 1 Given $\delta > 0$, we say that $t \mapsto \epsilon(t)$ satisfies the controlled decay property $(CD)_K$, if it is a nonincreasing function which satisfies: there exists $t_1 \geq t_0$ such that for all $t \geq t_1$,

$$\left(\frac{1}{\sqrt{\epsilon(t)}} \right)' \leq \min(2K - \delta, \delta - K),$$

where K is a parameter such that $\frac{\delta}{2} < K < \delta$ for $0 < \delta \leq 2$, and $\frac{\delta + \sqrt{\delta^2 - 4}}{2} < K < \delta$ for $\delta > 2$.

Theorem 5 Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of (TRIGS). Let δ be a positive parameter. Suppose that $\epsilon(\cdot)$ satisfies the condition $(CD)_K$ for some $K > 0$. Then, we have the following rate of convergence of values: for all $t \geq t_1$

$$f(x(t)) - \min_{\mathcal{H}} f \leq \frac{K \|x^*\|^2}{2} \frac{1}{\mathfrak{M}(t)} \int_{t_1}^t \epsilon^{\frac{3}{2}}(s) \mathfrak{M}(s) ds + \frac{C}{\mathfrak{M}(t)}, \quad (11)$$

where

$$\mathfrak{M}(t) = \exp \left(\int_{t_1}^t \mu(s) ds \right), \quad \mu(t) = -\frac{\dot{\epsilon}(t)}{2\epsilon(t)} + (\delta - K) \sqrt{\epsilon(t)}$$

and

$$C = (f(x(t_1)) - f(x^*)) + \frac{\epsilon(t_1)}{2} \|x(t_1)\|^2 + \frac{1}{2} \|K \sqrt{\epsilon(t_1)}(x(t_1) - x^*) + \dot{x}(t_1)\|^2.$$

Proof Lyapunov analysis. Set $f^* := f(x^*) = \min_{\mathcal{H}} f$. The energy function $\mathcal{E} : [t_0, +\infty[\rightarrow \mathbb{R}_+$,

$$\mathcal{E}(t) := (f(x(t)) - f^*) + \frac{\epsilon(t)}{2} \|x(t)\|^2 + \frac{1}{2} \|c(t)(x(t) - x^*) + \dot{x}(t)\|^2, \quad (12)$$

will be the basis for our Lyapunov analysis. The function $c : [t_0, +\infty[\rightarrow \mathbb{R}$ will be defined later, appropriately. Let us differentiate $\mathcal{E}(\cdot)$. By using the derivation chain rule, we get

$$\begin{aligned} \dot{\mathcal{E}}(t) &= \langle \nabla f(x(t)), \dot{x}(t) \rangle + \frac{\dot{\epsilon}(t)}{2} \|x(t)\|^2 + \epsilon(t) \langle \dot{x}(t), x(t) \rangle \\ &\quad + \langle c'(t)(x(t) - x^*) + c(t)\dot{x}(t) + \ddot{x}(t), c(t)(x(t) - x^*) + \dot{x}(t) \rangle. \end{aligned} \quad (13)$$

According to the constitutive equation (8), we have

$$\ddot{x}(t) = -\epsilon(t)x(t) - \delta \sqrt{\epsilon(t)} \dot{x}(t) - \nabla f(x(t)). \quad (14)$$

Therefore,

$$\begin{aligned} &\langle c'(t)(x(t) - x^*) + c(t)\dot{x}(t) + \ddot{x}(t), c(t)(x(t) - x^*) + \dot{x}(t) \rangle \\ &= \langle c'(t)(x(t) - x^*) + (c(t) - \delta \sqrt{\epsilon(t)}) \dot{x}(t) - (\epsilon(t)x(t) + \nabla f(x(t))), c(t)(x(t) - x^*) + \dot{x}(t) \rangle \\ &= c'(t)c(t) \|x(t) - x^*\|^2 + (c'(t) + c^2(t) - \delta c(t) \sqrt{\epsilon(t)}) \langle \dot{x}(t), x(t) - x^* \rangle + (c(t) - \delta \sqrt{\epsilon(t)}) \|\dot{x}(t)\|^2 \\ &\quad - \epsilon(t) \langle x(t), \dot{x}(t) \rangle - \langle \nabla f(x(t)), \dot{x}(t) \rangle - c(t) \langle \epsilon(t)x(t) + \nabla f(x(t)), x(t) - x^* \rangle. \end{aligned} \quad (15)$$

By combining (13) with (15), we get

$$\begin{aligned} \dot{\mathcal{E}}(t) &= \frac{\dot{\epsilon}(t)}{2} \|x(t)\|^2 + c'(t)c(t) \|x(t) - x^*\|^2 + (c'(t) + c^2(t) - \delta c(t) \sqrt{\epsilon(t)}) \langle \dot{x}(t), x(t) - x^* \rangle \\ &\quad + (c(t) - \delta \sqrt{\epsilon(t)}) \|\dot{x}(t)\|^2 - c(t) \langle \epsilon(t)x(t) + \nabla f(x(t)), x(t) - x^* \rangle. \end{aligned} \quad (16)$$

Consider the function

$$f_t : \mathcal{H} \longrightarrow \mathbb{R}, f_t(x) = f(x) + \frac{\epsilon(t)}{2} \|x\|^2.$$

According to the strong convexity property of f_t , we have

$$f_t(y) - f_t(x) \geq \langle \nabla f_t(x), y - x \rangle + \frac{\epsilon(t)}{2} \|x - y\|^2, \text{ for all } x, y \in \mathcal{H}.$$

Take $y = x^*$ and $x = x(t)$ in the above inequality. We get

$$\begin{aligned} f(x^*) + \frac{\epsilon(t)}{2} \|x^*\|^2 - f(x(t)) - \frac{\epsilon(t)}{2} \|x(t)\|^2 \geq \\ - \langle \nabla f(x(t)) + \epsilon(t)x(t), x(t) - x^* \rangle + \frac{\epsilon(t)}{2} \|x(t) - x^*\|^2. \end{aligned}$$

Consequently,

$$\begin{aligned} - \langle \nabla f(x(t)) + \epsilon(t)x(t), x(t) - x^* \rangle \leq - (f(x(t)) - f(x^*)) \\ + \frac{\epsilon(t)}{2} \|x^*\|^2 - \frac{\epsilon(t)}{2} \|x(t)\|^2 - \frac{\epsilon(t)}{2} \|x(t) - x^*\|^2. \end{aligned} \quad (17)$$

By multiplying (17) with $c(t)$ and injecting in (16) we get

$$\begin{aligned} \dot{\mathcal{E}}(t) \leq -c(t)(f(x(t)) - f^*) + \left(\frac{\dot{\epsilon}(t)}{2} - c(t) \frac{\epsilon(t)}{2} \right) \|x(t)\|^2 \\ + \left(c'(t)c(t) - c(t) \frac{\epsilon(t)}{2} \right) \|x(t) - x^*\|^2 + (c(t) - \delta\sqrt{\epsilon(t)}) \|\dot{x}(t)\|^2 \\ + (c'(t) + c^2(t) - \delta c(t)\sqrt{\epsilon(t)}) \langle \dot{x}(t), x(t) - x^* \rangle + c(t) \frac{\epsilon(t)}{2} \|x^*\|^2. \end{aligned} \quad (18)$$

On the other hand, for a positive function $\mu(t)$ we have

$$\begin{aligned} \mu(t)\mathcal{E}(t) = \mu(t)(f(x(t)) - f^*) + \mu(t) \frac{\epsilon(t)}{2} \|x(t)\|^2 + \frac{1}{2} \mu(t) c^2(t) \|x(t) - x^*\|^2 + \frac{1}{2} \mu(t) \|\dot{x}(t)\|^2 \\ + \mu(t) c(t) \langle \dot{x}(t), x(t) - x^* \rangle. \end{aligned} \quad (19)$$

By adding (18) and (19) we get

$$\begin{aligned} \dot{\mathcal{E}}(t) + \mu(t)\mathcal{E}(t) \leq (\mu(t) - c(t))(f(x(t)) - f^*) + \left(\frac{\dot{\epsilon}(t)}{2} - c(t) \frac{\epsilon(t)}{2} + \mu(t) \frac{\epsilon(t)}{2} \right) \|x(t)\|^2 \\ + \left(c'(t)c(t) - c(t) \frac{\epsilon(t)}{2} + \frac{1}{2} \mu(t) c^2(t) \right) \|x(t) - x^*\|^2 \\ + \left(c(t) - \delta\sqrt{\epsilon(t)} + \frac{1}{2} \mu(t) \right) \|\dot{x}(t)\|^2 \\ + \left(c'(t) + c^2(t) - \delta c(t)\sqrt{\epsilon(t)} + \mu(t)c(t) \right) \langle \dot{x}(t), x(t) - x^* \rangle + c(t) \frac{\epsilon(t)}{2} \|x^*\|^2. \end{aligned} \quad (20)$$

Since we have no control on the sign of $\langle \dot{x}(t), x(t) - x^* \rangle$, we take the coefficient in front of this term equal to zero, that is

$$c'(t) + c^2(t) - \delta c(t)\sqrt{\epsilon(t)} + \mu(t)c(t) = 0. \quad (21)$$

Take $c(t) = K\sqrt{\epsilon(t)}$. Indeed, it is here that the choice of c , and of the corresponding parameter K , come into play. The relation (21) can be equivalently written

$$\mu(t) = -\frac{\dot{\epsilon}(t)}{2\epsilon(t)} + (\delta - K)\sqrt{\epsilon(t)}.$$

According to this choice for $\mu(t)$ and $c(t)$, the inequality (20) becomes

$$\begin{aligned} \dot{\mathcal{E}}(t) + \mu(t)\mathcal{E}(t) &\leq \frac{1}{2\epsilon(t)} \left(-\dot{\epsilon}(t) + 2(\delta - 2K)\epsilon(t)^{\frac{3}{2}} \right) (f(x(t)) - f^*) \\ &\quad + \frac{1}{4} \left(\dot{\epsilon}(t) + 2(\delta - 2K)\epsilon(t)^{\frac{3}{2}} \right) \|x(t)\|^2 \\ &\quad + \frac{K}{4} \left(K\dot{\epsilon}(t) + 2\epsilon(t)^{\frac{3}{2}}(-K^2 + \delta K - 1) \right) \|x(t) - x^*\|^2 \\ &\quad + \frac{1}{4\epsilon(t)} \left(-\dot{\epsilon}(t) + 2(K - \delta)\epsilon(t)^{\frac{3}{2}} \right) \|\dot{x}(t)\|^2 + \frac{K\|x^*\|^2}{2} \epsilon^{\frac{3}{2}}(t). \end{aligned} \quad (22)$$

Let us show that the condition $(CD)_K$ provide the nonpositive sign for the coefficients in front of the terms of the right side of (22). Recall that, according to the hypotheses $(CD)_K$, for all $t \geq t_1$ we have the properties *a)* and *b)*:

$$\begin{aligned} a) \quad \left(\frac{1}{\sqrt{\epsilon(t)}} \right)' &\leq M_1(K) = \min(2K - \delta, \delta - K) = \begin{cases} 2K - \delta & \text{if } K \leq \frac{2}{3}\delta \\ \delta - K & \text{if } \frac{2}{3}\delta \leq K, \end{cases} \\ b) \quad \left(\frac{1}{\sqrt{\epsilon(t)}} \right)' &\geq 0. \end{aligned}$$

Without ambiguity we write briefly M_1 for $M_1(K)$. Note that *b)* just expresses that $\epsilon(\cdot)$ is non increasing. According to the hypotheses $(CD)_K$, we claim that for all $t \geq t_1$

$$\begin{cases} i) \quad \left(\frac{1}{\sqrt{\epsilon(t)}} \right)' \leq 2K - \delta \\ ii) \quad \left(\frac{1}{\sqrt{\epsilon(t)}} \right)' \geq \frac{\delta K - K^2 - 1}{K} \\ iii) \quad \left(\frac{1}{\sqrt{\epsilon(t)}} \right)' \leq \delta - K. \end{cases} \quad (23)$$

Let us justify these inequalities (23).

i) is a consequence of $\left(\frac{1}{\sqrt{\epsilon(t)}} \right)' \leq M_1$ and $M_1 \leq 2K - \delta$.

ii) is a consequence of $\left(\frac{1}{\sqrt{\epsilon(t)}} \right)' \geq 0$ and $\delta K - K^2 - 1 \leq 0$. Precisely, when $\delta \leq 2$ we have $\delta K - K^2 - 1 \leq 2K - K^2 - 1 \leq 0$. When $\delta > 2$, we have $\delta K - K^2 - 1 \leq 0$ because $K \geq \frac{\delta + \sqrt{\delta^2 - 4}}{2}$.

iii) is a consequence of $\left(\frac{1}{\sqrt{\epsilon(t)}} \right)' \leq M_1$ and $M_1 \leq \delta - K$.

The inequalities (23) can be equivalently written as follows: for all $t \geq t_1$

$$\begin{cases} i) \quad -\dot{\epsilon}(t) + 2(\delta - 2K)\epsilon(t)^{\frac{3}{2}} \leq 0 \\ ii) \quad K\dot{\epsilon}(t) + 2(\delta K - K^2 - 1)\epsilon(t)^{\frac{3}{2}} \leq 0 \\ iii) \quad -\dot{\epsilon}(t) + 2(K - \delta)\epsilon(t)^{\frac{3}{2}} \leq 0. \end{cases} \quad (24)$$

The inequalities (24) give that the coefficients entering the right side of (22) are nonpositive:

- *i)* gives that the coefficient of $f(x(t)) - f^*$ is nonpositive.
- Since $\dot{\epsilon}(t) \leq 0$ we have $\dot{\epsilon}(t) + 2(\delta - 2K)\epsilon(t)^{\frac{3}{2}} \leq -\dot{\epsilon}(t) + 2(\delta - 2K)\epsilon(t)^{\frac{3}{2}}$. Therefore, by *i)* we have that the coefficient of $\|x(t)\|^2$ in (22) is nonpositive.

- *ii*) gives that the coefficient of $\|x(t) - x^*\|^2$ is nonpositive.
- *iii*) gives that the coefficient of $\|\dot{x}(t)\|^2$ is nonpositive .

Let us return to (22). Using (24) and the above results, we obtain

$$\dot{\mathcal{E}}(t) + \mu(t)\mathcal{E}(t) \leq \frac{K\|x^*\|^2}{2}\epsilon^{\frac{3}{2}}(t), \text{ for all } t \geq t_1. \quad (25)$$

By multiplying (25) with $\mathfrak{M}(t) = \exp\left(\int_{t_1}^t \mu(s)ds\right)$ we obtain

$$\frac{d}{dt}(\mathfrak{M}(t)\mathcal{E}(t)) \leq \frac{K\|x^*\|^2}{2}\epsilon^{\frac{3}{2}}(t)\mathfrak{M}(t). \quad (26)$$

By integrating (26) on $[t_1, t]$ we get

$$\mathcal{E}(t) \leq \frac{K\|x^*\|^2}{2} \frac{\int_{t_1}^t \epsilon^{\frac{3}{2}}(s)\mathfrak{M}(s)ds}{\mathfrak{M}(t)} + \frac{\mathfrak{M}(t_1)\mathcal{E}(t_1)}{\mathfrak{M}(t)}. \quad (27)$$

By definition of $\mathcal{E}(t)$ we deduce that

$$f(x(t)) - \min_{\mathcal{H}} f \leq \frac{K\|x^*\|^2}{2} \frac{\int_{t_1}^t \epsilon^{\frac{3}{2}}(s)\mathfrak{M}(s)ds}{\mathfrak{M}(t)} + \frac{\mathcal{E}(t_1)}{\mathfrak{M}(t)}, \quad (28)$$

for all $t \geq t_1$, and this gives the convergence rate of the values.

Remark 1 By integrating the relation $0 \leq \left(\frac{1}{\sqrt{\epsilon(t)}}\right)' \leq M_1$ on an interval $[t_1, t]$, we get

$$\frac{1}{\sqrt{\epsilon(t_1)}} \leq \frac{1}{\sqrt{\epsilon(t)}} \leq M_1 t + \frac{1}{\sqrt{\epsilon(t_1)}} - M_1 t_1.$$

Therefore, denoting $C_1 = \frac{1}{\sqrt{\epsilon(t_1)}} - M_1 t_1$, and $C_2 = \epsilon(t_1)$ we have

$$\frac{1}{(M_1 t + C_1)^2} \leq \epsilon(t) \leq C_2. \quad (29)$$

This shows that the Lyapunov analysis developed previously only provides information in the case where $\epsilon(t)$ is greater than or equal to C/t^2 . Since the damping coefficient $\gamma(t) = \delta\sqrt{\epsilon(t)}$, this means that $\gamma(t)$ must be greater than or equal to C/t . This is in accordance with the theory of inertial gradient systems with time-dependent viscosity coefficient, which states that the asymptotic optimization property is valid provided that the integral on $[t_0, +\infty[$ of $\gamma(t)$ is infinite, see [8].

As a consequence of Theorem 5 we have the following result.

Corollary 1 *Under the hypothesis of Theorem 5 we have*

$$\lim_{t \rightarrow +\infty} \mathfrak{M}(t) = +\infty. \quad (30)$$

Suppose moreover that $\epsilon^{\frac{3}{2}}(\cdot) \in L^1(t_0, +\infty)$. Then

$$\lim_{t \rightarrow +\infty} f(x(t)) = \min_{\mathcal{H}} f. \quad (31)$$

Proof By definition of $\mu(t)$, since $\epsilon(\cdot)$ is nonincreasing and $\delta \geq K$, we have that $\mu(t)$ is nonnegative for all $t \geq t_1$. Therefore, $t \mapsto \mathfrak{M}(t)$ is a nondecreasing function. Let us write equivalently $\mu(t) = \frac{d}{dt} \ln \frac{1}{\sqrt{\epsilon(t)}} + (\delta - K)\sqrt{\epsilon(t)}$, and integrate on $[t_1, t]$. We obtain

$$\mathfrak{M}(t) = \exp\left(\int_{t_1}^t \mu(s) ds\right) = \frac{C}{\sqrt{\epsilon(t)}} \exp\left(\int_{t_1}^t (\delta - K)\sqrt{\epsilon(s)} ds\right).$$

Since $\delta - K \geq 0$, we deduce that $\mathfrak{M}(t) \geq \frac{C}{\sqrt{\epsilon(t)}}$. Since $\lim_{t \rightarrow \infty} \epsilon(t) = 0$, we get

$$\lim_{t \rightarrow +\infty} \mathfrak{M}(t) = +\infty.$$

Moreover, if we suppose that $\epsilon^{\frac{3}{2}}(\cdot) \in L^1(t_0, +\infty)$, then by [11, Lemma A.3] we obtain

$$\lim_{t \rightarrow +\infty} \frac{\int_{t_1}^t \epsilon^{\frac{3}{2}}(s) \mathfrak{M}(s) ds}{\mathfrak{M}(t)} = 0.$$

Combining these properties with the convergence rate (11) of Theorem 5, we obtain (31).

2.3 Particular cases

Since $\epsilon(t) \rightarrow 0$ as $t \rightarrow +\infty$, (TRIGS) falls within the setting of the inertial dynamics with an asymptotic vanishing damping coefficient $\gamma(t)$. Here, $\gamma(t) = \delta\sqrt{\epsilon(t)}$. We know with Cabot-Engler-Gaddat [27] that for such systems, the optimization property is satisfied asymptotically if $\int_{t_0}^{+\infty} \gamma(t) dt = +\infty$ (i.e. $\gamma(t)$ does not tend too rapidly towards zero). By taking $\epsilon(t) = \frac{c}{t^p}$, it is easy to verify that the condition $(CD)_K$ is satisfied if $p \leq 2$, that is $\sqrt{\epsilon(t)} = \frac{c}{t^p}$, with $p \leq 1$, which is in accordance with the above property. Let us particularize Theorem 5 to situations where the integrals can be computed (at least estimated).

2.3.1 $\epsilon(t)$ of order $1/t^2$

Take

$$\epsilon(t) = \frac{1}{(Mt + C)^2}, \quad M < M_1(K), \quad C \leq C_1.$$

Then, $\left(\frac{1}{\sqrt{\epsilon(t)}}\right)' \leq M_1(K)$ for all $t \geq t_0$ and the condition $(CD)_K$ is satisfied. Moreover,

$$\mu(t) = \frac{M + \delta - K}{Mt + C}, \quad \mathfrak{M}(t) = \left(\frac{Mt + C}{Mt_0 + C}\right)^{\frac{M + \delta - K}{M}}.$$

Therefore, (11) becomes

$$\mathcal{E}(t) \leq \frac{K\|x^*\|^2}{2} \frac{\int_{t_0}^t (Ms + C)^{\frac{-2M + \delta - K}{M}} ds}{(Mt + C)^{\frac{M + \delta - K}{M}}} + \frac{(Mt_0 + C)^{\frac{M + \delta - K}{M}} \mathcal{E}(t_0)}{(Mt + C)^{\frac{M + \delta - K}{M}}}. \quad (32)$$

Consequently, we have

$$\mathcal{E}(t) \leq \frac{K\|x^*\|^2}{2(-M + \delta - K)} \frac{1}{(Mt + C)^2} + \frac{-\frac{K\|x^*\|^2}{2(-M + \delta - K)} (Mt_0 + C)^{\frac{-M + \delta - K}{M}} + (Mt_0 + C)^{\frac{M + \delta - K}{M}} \mathcal{E}(t_0)}{(Mt + C)^{\frac{M + \delta - K}{M}}}.$$

By assumption we have $M < M_1 \leq \delta - K$. Therefore $\frac{M+\delta-K}{M} > 2$ and $-M + \delta - K > 0$. It follows that when $Mt + C \geq 1$

$$\mathcal{E}(t) \leq \frac{C'}{(Mt + C)^2}, \quad \text{with } C' = \frac{K\|x^*\|^2}{2(-M + \delta - K)} + (Mt_0 + C)^{\frac{M+\delta-K}{M}} \mathcal{E}(t_0).$$

Observe that $\delta\sqrt{\epsilon(t)} = \frac{\frac{\delta}{M}}{t + \frac{C}{M}} = \frac{\alpha}{t + \beta}$, where we set $\alpha = \frac{\delta}{M}$ and $\beta = \frac{C}{M}$. Since $M < M_1 \leq \frac{1}{3}\delta$ we get $\alpha \in]3, +\infty[$. Indeed, we can get any $\alpha > 3$. Note also that by translating the time scale the result in the general case $\beta \geq 0$ results from its obtaining for a particular case $\beta = 0$. According to the fact that we can take for δ any positive number, we obtain

Theorem 6 Take $\alpha \in]3, +\infty[$, $c > 0$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) + \frac{c}{t^2}x(t) = 0.$$

Then, the following convergence rate of the values is satisfied: as $t \rightarrow +\infty$

$$f(x(t)) - \min_{\mathcal{H}} f = O\left(\frac{1}{t^2}\right).$$

Remark 2 It is a natural question to compare our dynamic ($c > 0$) with the Su-Boyd-Candès dynamic [38] ($c = 0$), which was introduced as a continuous version of the Nesterov accelerated gradient method. We obtain the optimal convergence rate of values with an additional Tikhonov regularization term, which is a remarkable property. In fact, in the next sections we will prove that the Tikhonov term induces strong convergence of the trajectory to the minimum norm solution.

2.3.2 $\epsilon(t)$ of order $1/t^r$, $\frac{2}{3} < r < 2$

Take $\epsilon(t) = 1/t^r$, $r < 2$. Then

$$\begin{aligned} \mu(t) &= -\frac{1}{2} \frac{\dot{\epsilon}(t)}{\epsilon(t)} + (\delta - K)\sqrt{\epsilon(t)} \\ &= \frac{r}{2t} + \frac{\delta - K}{t^{\frac{r}{2}}}. \end{aligned}$$

Therefore

$$\mathfrak{M}(t) = \exp \int_{t_0}^t \left(\frac{r}{2s} + \frac{\delta - K}{s^{\frac{r}{2}}} \right) ds = Ct^{\frac{r}{2}} \exp \left(\frac{2(\delta - K)}{2 - r} t^{1 - \frac{r}{2}} \right).$$

Set

$$m(t) := t^{\frac{r}{2}} \exp \left(\frac{2(\delta - K)}{2 - r} t^{1 - \frac{r}{2}} \right).$$

According to (28) we have that for some $C_1 > 0$

$$f(x(t)) - \min_{\mathcal{H}} f \leq \frac{C_1}{m(t)} \int_{t_0}^t \frac{m(s)}{s^{\frac{3r}{2}}} ds + \frac{C_1}{m(t)}. \quad (33)$$

Note that according to $r < 2$, $m(t)$ is an increasing function which has an exponential growth as $t \rightarrow +\infty$. Accordingly, by the mean value theorem we have the following majorization.

$$\frac{1}{m(t)} \int_{t_0}^t \frac{m(s)}{s^{\frac{3r}{2}}} ds \leq \frac{m(t)}{m(t)} \int_{t_0}^t \frac{1}{s^{\frac{3r}{2}}} ds = \mathcal{O} \left(\frac{1}{t^{\frac{3r}{2} - 1}} \right). \quad (34)$$

Let us summarize these results in the following statement.

Theorem 7 Take $\epsilon(t) = 1/t^r$, $\frac{2}{3} < r < 2$, $\delta > 0$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a global trajectory of

$$\ddot{x}(t) + \frac{\delta}{t^{\frac{r}{2}}}\dot{x}(t) + \nabla f(x(t)) + \frac{1}{t^r}x(t) = 0.$$

Then, the following convergence rate of the values is satisfied: as $t \rightarrow +\infty$

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{\frac{3r}{2}-1}}\right).$$

Remark 3 When $r \rightarrow 2$ the exponent $\frac{3r}{2} - 1$ tends to 2. So there is a continuous transition in the convergence rate. As in Remark 2 the additional Tikhonov regularization term is expected to have a regularization effect (even better than in the case $r = 2$). In addition, the above analysis makes appear another critical value, namely $r = \frac{2}{3}$.

3 In-depth analysis in the critical case $\epsilon(t) = c/t^2$

Let us refine our analysis in the case where the Tikhonov regularization coefficient and the damping coefficient are respectively of order $1/t^2$ and $1/t$. Our analysis will now take into account the coefficients α and c in front of these terms. So the Cauchy problem for (TRIGS) is written

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) + \frac{c}{t^2}x(t) = 0 \\ x(t_0) = x_0, \dot{x}(t_0) = v_0, \end{cases} \quad (35)$$

where $t_0 > 0$, $c > 0$, $(x_0, v_0) \in \mathcal{H} \times \mathcal{H}$, and $\alpha \geq 3$. The starting time t_0 is taken strictly greater than zero to take into account the fact that the functions $\frac{c}{t^2}$ and $\frac{\alpha}{t}$ have singularities at 0. This is not a limitation of the generality of the proposed approach, since we will focus on the asymptotic behaviour of the generated trajectories.

3.1 Convergence rate of the values

Theorem 8 Let $t_0 > 0$ and, for some initial data $x_0, v_0 \in \mathcal{H}$, let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be the unique global solution of (35). Then, the following results hold.

- i) If $\alpha = 3$, then $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{\ln t}{t^2}\right)$ as $t \rightarrow +\infty$.
- ii) If $\alpha > 3$, then $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right)$ as $t \rightarrow +\infty$. Further, the trajectory x is bounded and $\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right)$ as $t \rightarrow +\infty$.

Proof The analysis is parallel to that of Theorem 5. Set $f^* := f(x^*) = \min_{\mathcal{H}} f$. Let $b : [t_0, +\infty[\rightarrow \mathbb{R}$, $b(t) = \frac{K}{t}$ where $K > 0$ will be defined later. Let us introduce $\mathcal{E} : [t_0, +\infty[\rightarrow \mathbb{R}$,

$$\mathcal{E}(t) := (f(x(t)) - f^*) + \frac{c}{2t^2}\|x(t)\|^2 + \frac{1}{2}\|b(t)(x(t) - x^*) + \dot{x}(t)\|^2, \quad (36)$$

that will serve as a Lyapunov function. Then,

$$\begin{aligned} \dot{\mathcal{E}}(t) &= \langle \nabla f(x(t)), \dot{x}(t) \rangle - \frac{c}{t^3}\|x(t)\|^2 + \frac{c}{t^2}\langle \dot{x}(t), x(t) \rangle \\ &\quad + \langle b'(t)(x(t) - x^*) + b(t)\dot{x}(t) + \ddot{x}(t), b(t)(x(t) - x^*) + \dot{x}(t) \rangle. \end{aligned} \quad (37)$$

According to the dynamic system (35), we have

$$\ddot{x}(t) = -\frac{c}{t^2}x(t) - \frac{\alpha}{t}\dot{x}(t) - \nabla f(x(t)). \quad (38)$$

Therefore,

$$\begin{aligned} & \langle b'(t)(x(t) - x^*) + b(t)\dot{x}(t) + \ddot{x}(t), b(t)(x(t) - x^*) + \dot{x}(t) \rangle = \\ & \left\langle -\frac{K}{t^2}(x(t) - x^*) + \frac{K - \alpha}{t}\dot{x}(t) - \left(\frac{c}{t^2}x(t) + \nabla f(x(t))\right), \frac{K}{t}(x(t) - x^*) + \dot{x}(t) \right\rangle = \\ & -\frac{K^2}{t^3}\|x(t) - x^*\|^2 + \frac{K^2 - \alpha K - K}{t^2}\langle \dot{x}(t), x(t) - x^* \rangle + \frac{K - \alpha}{t}\|\dot{x}(t)\|^2 \\ & -\frac{c}{t^2}\langle x(t), \dot{x}(t) \rangle - \langle \nabla f(x(t)), \dot{x}(t) \rangle - \frac{K}{t}\left\langle \frac{c}{t^2}x(t) + \nabla f(x(t)), x(t) - x^* \right\rangle. \end{aligned} \quad (39)$$

Combining (37) and (39), we get

$$\begin{aligned} \dot{\mathcal{E}}(t) = & -\frac{c}{t^3}\|x(t)\|^2 - \frac{K^2}{t^3}\|x(t) - x^*\|^2 + \frac{K^2 - \alpha K - K}{t^2}\langle \dot{x}(t), x(t) - x^* \rangle + \frac{K - \alpha}{t}\|\dot{x}(t)\|^2 \\ & - \frac{K}{t}\left\langle \frac{c}{t^2}x(t) + \nabla f(x(t)), x(t) - x^* \right\rangle. \end{aligned} \quad (40)$$

Consider the strongly convex function

$$f_t : \mathcal{H} \longrightarrow \mathbb{R}, \quad f_t(x) = f(x) + \frac{c}{2t^2}\|x\|^2.$$

From the gradient inequality we have

$$f_t(y) - f_t(x) \geq \langle \nabla f_t(x), y - x \rangle + \frac{c}{2t^2}\|x - y\|^2, \quad \text{for all } x, y \in \mathcal{H}.$$

Take $y = x^*$ and $x = x(t)$ in the above inequality. We obtain

$$\begin{aligned} f^* + \frac{c}{2t^2}\|x^*\|^2 - f(x(t)) - \frac{c}{2t^2}\|x(t)\|^2 \geq \\ -\left\langle \nabla f(x(t)) + \frac{c}{t^2}x(t), x(t) - x^* \right\rangle + \frac{c}{2t^2}\|x(t) - x^*\|^2. \end{aligned}$$

Consequently,

$$\begin{aligned} -\left\langle \frac{c}{t^2}x(t) + \nabla f(x(t)), x(t) - x^* \right\rangle \leq & -(f(x(t)) - f^*) - \frac{c}{2t^2}\|x(t)\|^2 - \frac{c}{2t^2}\|x(t) - x^*\|^2 \\ & + \frac{c}{2t^2}\|x^*\|^2. \end{aligned} \quad (41)$$

By multiplying (41) with $\frac{K}{t}$, and injecting in (40), we obtain

$$\begin{aligned} \dot{\mathcal{E}}(t) \leq & -\frac{K}{t}(f(x(t)) - f^*) - \left(\frac{c}{t^3} + \frac{Kc}{2t^3}\right)\|x(t)\|^2 - \left(\frac{K^2}{t^3} + \frac{Kc}{2t^3}\right)\|x(t) - x^*\|^2 \\ & + \frac{K^2 - \alpha K - K}{t^2}\langle \dot{x}(t), x(t) - x^* \rangle + \frac{K - \alpha}{t}\|\dot{x}(t)\|^2 + \frac{cK}{2t^3}\|x^*\|^2. \end{aligned} \quad (42)$$

On the other hand, by multiplying the function $\mathcal{E}(t)$ by $\mu(t) = \frac{\alpha - K + 1}{t}$, we obtain

$$\begin{aligned} \mu(t)\mathcal{E}(t) = & \frac{\alpha - K + 1}{t}(f(x(t)) - f^*) + \frac{(\alpha - K + 1)c}{2t^3}\|x(t)\|^2 + \frac{(\alpha - K + 1)K^2}{2t^3}\|x(t) - x^*\|^2 \\ & + \frac{\alpha - K + 1}{2t}\|\dot{x}(t)\|^2 + \frac{(\alpha - K + 1)K}{t^2}\langle \dot{x}(t), x(t) - x^* \rangle. \end{aligned} \quad (43)$$

By adding (42) and (43), we get

$$\begin{aligned} \dot{\mathcal{E}}(t) + \mu(t)\mathcal{E}(t) &\leq \frac{\alpha - 2K + 1}{t} (f(x(t)) - f^*) + \frac{(\alpha - 2K - 1)c}{2t^3} \|x(t)\|^2 \\ &\quad + \frac{(\alpha - K - 1)K^2 - Kc}{2t^3} \|x(t) - x^*\|^2 + \frac{K - \alpha + 1}{2t} \|\dot{x}(t)\|^2 + \frac{cK}{2t^3} \|x^*\|^2. \end{aligned} \quad (44)$$

The case $\alpha > 3$. Take $\frac{\alpha+1}{2} < K < \alpha - 1$. Since $\alpha > 3$, such K exists. This implies that $\alpha - 2K + 1 < 0$, hence $\alpha - 2K - 1 < 0$, and $K - \alpha + 1 < 0$. In addition, since $c > 0$ there exists $K \in]\frac{\alpha+1}{2}, \alpha - 1[$ such that

$$(\alpha - K - 1)K^2 - Kc \leq 0. \quad (45)$$

Indeed, (45) can be deduced from the fact that the continuous function $\varphi(K) = (\alpha - K - 1)K$ is decreasing on the interval $[\frac{\alpha+1}{2}, \alpha - 1]$ and $\varphi(\alpha - 1) = 0$. Therefore, for every $c > 0$ there exists $K \in]\frac{\alpha+1}{2}, \alpha - 1[$ such that $c \geq \varphi(K)$. So take $K \in]\frac{\alpha+1}{2}, \alpha - 1[$ such that (45) holds. Then, by collecting the previous results, (44) yields

$$\dot{\mathcal{E}}(t) + \mu(t)\mathcal{E}(t) \leq \frac{cK}{2t^3} \|x^*\|^2. \quad (46)$$

Taking into account that $\mu(t) = \frac{\alpha - K + 1}{t}$, by multiplying (46) with $t^{\alpha - K + 1}$ we get

$$\frac{d}{dt} \left(t^{\alpha - K + 1} \mathcal{E}(t) \right) \leq \frac{cK}{2} \|x^*\|^2 t^{\alpha - K - 2}. \quad (47)$$

By integrating (47) on $[t_0, t]$, we get

$$\mathcal{E}(t) \leq \frac{cK \|x^*\|^2}{2(\alpha - K - 1)} \frac{1}{t^2} - \frac{cK \|x^*\|^2}{2(\alpha - K - 1)} \frac{t_0^{\alpha - K - 1}}{t^{\alpha - K + 1}} + \frac{t_0^{\alpha - K + 1} \mathcal{E}(t_0)}{t^{\alpha - K + 1}}. \quad (48)$$

Since $\alpha - K + 1 > 2$, we obtain

$$\mathcal{E}(t) = O\left(\frac{1}{t^2}\right) \text{ as } t \rightarrow +\infty. \quad (49)$$

By definition of $\mathcal{E}(t)$ we immediately deduce that

$$f(x(t)) - \min_{\mathcal{H}} f = O\left(\frac{1}{t^2}\right) \text{ as } t \rightarrow +\infty, \quad (50)$$

and further, that the trajectory $x(\cdot)$ is bounded and

$$\|\dot{x}(t)\| = O\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty.$$

The case $\alpha = 3$. Take $K = 2$. With the previous notations, we have now $\mu(t) = \frac{2}{t}$ and (44) gives

$$\dot{\mathcal{E}}(t) + \frac{2}{t}\mathcal{E}(t) \leq -\frac{c}{t^3} \|x(t)\|^2 - \frac{c}{t^3} \|x(t) - x^*\|^2 + \frac{c}{t^3} \|x^*\|^2 \leq \frac{c}{t^3} \|x^*\|^2. \quad (51)$$

After multiplication of (51) by t^2 we get

$$\frac{d}{dt} (t^2 \mathcal{E}(t)) \leq \frac{c}{t} \|x^*\|^2. \quad (52)$$

By integrating (52) on $[t_0, t]$ we get

$$\mathcal{E}(t) \leq c \|x^*\|^2 \frac{\ln t}{t^2} - c \|x^*\|^2 \frac{\ln t_0}{t^2} + \frac{t_0^2 \mathcal{E}(t_0)}{t^2}. \quad (53)$$

Consequently, we have

$$\mathcal{E}(t) = O\left(\frac{\ln t}{t^2}\right) \text{ as } t \rightarrow +\infty. \quad (54)$$

By definition of $\mathcal{E}(t)$ we immediately deduce that

$$f(x(t)) - \min f = O\left(\frac{\ln t}{t^2}\right) \text{ as } t \rightarrow +\infty. \quad (55)$$

which gives the claim.

3.2 Strong convergence

Theorem 9 *Let $t_0 > 0$ and, for some starting points $x_0, v_0 \in \mathcal{H}$, let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be the unique global solution of (35). Let x^* be the element of minimal norm of $S = \operatorname{argmin} f$, that is $x^* = \operatorname{proj}_S 0$. Then, for all $\alpha > 3$ we have that*

$$\liminf_{t \rightarrow +\infty} \|x(t) - x^*\| = 0.$$

Further, if there exists $T \geq t_0$, such that the trajectory $\{x(t) : t \geq T\}$ stays either in the open ball $B(0, \|x^\|)$ or in its complement, then $x(t)$ converges strongly to x^* as $t \rightarrow +\infty$.*

Proof The proof combines energetic and geometric arguments, as it was initiated in [13]. We successively consider the three following configurations of the trajectory.

I. Assume that there exists $T \geq t_0$ such that $\|x(t)\| \geq \|x^*\|$ for all $t \geq T$. Let us denote $f_t(x) := f(x) + \frac{c}{2t^2}\|x\|^2$ and let $x_t := \operatorname{argmin} f_t(x)$. Let us recall some classical properties of the Tikhonov approximation:

$$\forall t > 0 \quad \|x_t\| \leq \|x^*\|, \quad \text{and} \quad \lim_{t \rightarrow +\infty} \|x_t - x^*\| = 0. \quad (56)$$

Using the gradient inequality for the strongly convex function f_t , we have

$$f_t(x(t)) - f_t(x_t) \geq \frac{c}{2t^2} \|x(t) - x_t\|^2.$$

On the other hand

$$f_t(x_t) - f_t(x^*) = f(x_t) - f^* + \frac{c}{2t^2} (\|x_t\|^2 - \|x^*\|^2) \geq \frac{c}{2t^2} (\|x_t\|^2 - \|x^*\|^2).$$

By adding the last two inequalities we get

$$f_t(x(t)) - f_t(x^*) \geq \frac{c}{2t^2} (\|x(t) - x_t\|^2 + \|x_t\|^2 - \|x^*\|^2), \quad (57)$$

Therefore, according to (56), to obtain the strong convergence of the trajectory $x(t)$ to x^* , it is enough to show that $f_t(x(t)) - f_t(x^*) = o\left(\frac{1}{t^2}\right)$, as $t \rightarrow +\infty$.

For $K > 0$, consider now the energy functional

$$\begin{aligned} E(t) &= f_t(x(t)) - f_t(x^*) + \frac{1}{2} \left\| \frac{K}{t} (x(t) - x^*) + \dot{x}(t) \right\|^2 \\ &= (f(x(t)) - f(x^*)) + \frac{c}{2t^2} (\|x(t)\|^2 - \|x^*\|^2) + \frac{1}{2} \left\| \frac{K}{t} (x(t) - x^*) + \dot{x}(t) \right\|^2. \end{aligned} \quad (58)$$

Then,

$$\begin{aligned} \dot{E}(t) = & \langle \nabla f_t(x(t)), \dot{x}(t) \rangle - \frac{c}{2t^3} (\|x(t)\|^2 - \|x^*\|^2) \\ & + \left\langle -\frac{K}{t^2}(x(t) - x^*) + \frac{K}{t}\dot{x}(t) + \ddot{x}(t), \frac{K}{t}(x(t) - x^*) + \dot{x}(t) \right\rangle. \end{aligned} \quad (59)$$

Let us examine the different terms of (59). According to the constitutive equation (35) we have

$$\begin{aligned} & \left\langle -\frac{K}{t^2}(x(t) - x^*) + \frac{K}{t}\dot{x}(t) + \ddot{x}(t), \frac{K}{t}(x(t) - x^*) + \dot{x}(t) \right\rangle = \\ & \left\langle -\frac{K}{t^2}(x(t) - x^*) + \frac{K - \alpha}{t}\dot{x}(t) - \left(\frac{c}{t^2}x(t) + \nabla f(x(t)) \right), \frac{K}{t}(x(t) - x^*) + \dot{x}(t) \right\rangle = \\ & -\frac{K^2}{t^3} \|x(t) - x^*\|^2 + \frac{K^2 - \alpha K - K}{t^2} \langle \dot{x}(t), x(t) - x^* \rangle + \frac{K - \alpha}{t} \|\dot{x}(t)\|^2 \\ & - \frac{c}{t^2} \langle x(t), \dot{x}(t) \rangle - \langle \nabla f(x(t)), \dot{x}(t) \rangle - \frac{K}{t} \left\langle \frac{c}{t^2}x(t) + \nabla f(x(t)), x(t) - x^* \right\rangle. \end{aligned} \quad (60)$$

Further, from (41) we get

$$\begin{aligned} & -\frac{K}{t} \left\langle \frac{c}{t^2}x(t) + \nabla f(x(t)), x(t) - x^* \right\rangle \\ & \leq -\frac{K}{t} (f(x(t)) - f^*) - \frac{cK}{2t^3} \|x(t)\|^2 - \frac{cK}{2t^3} \|x(t) - x^*\|^2 + \frac{cK}{2t^3} \|x^*\|^2 \\ & = -\frac{K}{t} (f_t(x(t)) - f_t(x^*)) - \frac{cK}{2t^3} \|x(t) - x^*\|^2. \end{aligned} \quad (61)$$

Injecting (60) and (61) in (59) we get

$$\begin{aligned} \dot{E}(t) \leq & -\frac{K}{t} (f_t(x(t)) - f_t(x^*)) - \frac{c}{t^3} (\|x(t)\|^2 - \|x^*\|^2) - \frac{2K^2 + cK}{2t^3} \|x(t) - x^*\|^2 \\ & + \frac{K^2 - \alpha K - K}{t^2} \langle \dot{x}(t), x(t) - x^* \rangle + \frac{K - \alpha}{t} \|\dot{x}(t)\|^2. \end{aligned} \quad (62)$$

Consider now the function $\mu(t) = \frac{\alpha+1-K}{t}$. Then,

$$\begin{aligned} \mu(t)E(t) = & \frac{\alpha+1-K}{t} (f_t(x(t)) - f_t(x^*)) + \frac{K^2(\alpha+1-K)}{2t^3} \|x(t) - x^*\|^2 \\ & + \frac{K(\alpha+1-K)}{t^2} \langle \dot{x}(t), x(t) - x^* \rangle + \frac{\alpha+1-K}{2t} \|\dot{x}(t)\|^2. \end{aligned} \quad (63)$$

Consequently, (62) and (63) yield

$$\begin{aligned} \dot{E}(t) + \mu(t)E(t) \leq & \frac{\alpha+1-2K}{t} (f_t(x(t)) - f_t(x^*)) - \frac{c}{t^3} (\|x(t)\|^2 - \|x^*\|^2) \\ & + \frac{K^2(\alpha-1-K) - cK}{2t^3} \|x(t) - x^*\|^2 + \frac{K - \alpha + 1}{2t} \|\dot{x}(t)\|^2 \\ = & \frac{\alpha+1-2K}{t} (f(x(t)) - f(x^*)) + (\alpha-1-2K) \frac{c}{2t^3} (\|x(t)\|^2 - \|x^*\|^2) \\ & + \frac{K^2(\alpha-1-K) - cK}{2t^3} \|x(t) - x^*\|^2 + \frac{K - \alpha + 1}{2t} \|\dot{x}(t)\|^2. \end{aligned} \quad (64)$$

Assume that $\frac{\alpha+1}{2} < K < \alpha - 1$. Since $\alpha > 3$ such K exists. As in the proof of Theorem 8 we deduce that $\alpha - 2K + 1 < 0$, $K - \alpha + 1 < 0$ and since $c > 0$ there exists $K \in (\frac{\alpha+1}{2}, \alpha - 1)$ such that

$$(\alpha - K - 1)K^2 - Kc \leq 0. \quad (65)$$

So take $K \in (\frac{\alpha+1}{2}, \alpha - 1)$ such that (65) holds. Then, (64) leads to

$$\dot{E}(t) + \frac{\alpha + 1 - K}{t} E(t) \leq (\alpha - 1 - 2K) \frac{c}{2t^3} (\|x(t)\|^2 - \|x^*\|^2). \quad (66)$$

Let us integrate the differential inequality (66). After multiplication by $t^{\alpha+1-K}$ we get

$$\frac{d}{dt} t^{\alpha+1-K} E(t) \leq \frac{c}{2} (\alpha - 1 - 2K) t^{\alpha-2-K} (\|x(t)\|^2 - \|x^*\|^2)$$

and integrating the latter on $[T, t]$, $t > T$ we obtain

$$E(t) \leq \frac{c}{2} (\alpha - 1 - 2K) \frac{\int_T^t s^{\alpha-2-K} (\|x(s)\|^2 - \|x^*\|^2) ds}{t^{\alpha+1-K}} + \frac{T^{\alpha+1-K} E(T)}{t^{\alpha+1-K}}. \quad (67)$$

In one hand, from the definition of $E(t)$ we have

$$f_t(x(t)) - f_t(x^*) \leq E(t).$$

Therefore,

$$f_t(x(t)) - f_t(x^*) \leq \frac{c}{2} (\alpha - 1 - 2K) \frac{\int_T^t s^{\alpha-2-K} (\|x(s)\|^2 - \|x^*\|^2) ds}{t^{\alpha+1-K}} + \frac{T^{\alpha+1-K} E(T)}{t^{\alpha+1-K}}.$$

On the other hand (57) gives

$$f_t(x(t)) - f_t(x^*) \geq \frac{c}{2t^2} (\|x(t) - x_t\|^2 + \|x_t\|^2 - \|x^*\|^2).$$

Consequently,

$$(\alpha - 1 - 2K) \frac{\int_T^t s^{\alpha-2-K} (\|x(s)\|^2 - \|x^*\|^2) ds}{t^{\alpha-1-K}} + \frac{2T^{\alpha+1-K} E(T)}{ct^{\alpha-1-K}} \geq \|x(t) - x_t\|^2 + \|x_t\|^2 - \|x^*\|^2. \quad (68)$$

By assumption $\|x(t)\| \geq \|x^*\|$ for all $t \geq T$ and $\alpha - 1 - 2K < 0$. Hence, for all $t > T$, (68) leads to

$$\frac{2T^{\alpha+1-K} E(T)}{ct^{\alpha-1-K}} \geq \|x(t) - x_t\|^2 + \|x_t\|^2 - \|x^*\|^2. \quad (69)$$

Now, by taking the limit $t \rightarrow +\infty$ and using that $x_t \rightarrow x^*$, $t \rightarrow +\infty$ we get

$$\lim_{t \rightarrow +\infty} \|x(t) - x_t\| \leq 0$$

and hence

$$\lim_{t \rightarrow +\infty} x(t) = x^*.$$

II. Assume now that there exists $T \geq t_0$ such that $\|x(t)\| < \|x^*\|$ for all $t \geq T$. According to Theorem 8, we have that

$$\lim_{t \rightarrow +\infty} f(x(t)) = \min_{\mathcal{H}} f.$$

Let $\bar{x} \in \mathcal{H}$ be a weak sequential cluster point of the trajectory x , which exists since, by Theorem 8, the trajectory is bounded. So, there exists a sequence $(t_n)_{n \in \mathbb{N}} \subseteq [T, +\infty)$ such that $t_n \rightarrow +\infty$ and $x(t_n)$ converges weakly to \bar{x} as $n \rightarrow +\infty$. Since f is weakly lower semicontinuous, we deduce that

$$f(\bar{x}) \leq \liminf_{n \rightarrow +\infty} f(x(t_n)) = \min_{\mathcal{H}} f,$$

hence $\bar{x} \in \operatorname{argmin} f$. Now, since the norm is weakly lower semicontinuous, and since $\|x(t)\| < \|x^*\|$ for all $t \geq T$, we have

$$\|\bar{x}\| \leq \liminf_{n \rightarrow +\infty} \|x(t_n)\| \leq \|x^*\|.$$

Combining $\bar{x} \in \operatorname{argmin} f$ with the definition of x^* , this implies that $\bar{x} = x^*$. This shows that the trajectory $x(\cdot)$ converges weakly to x^* . So

$$\|x^*\| \leq \liminf_{t \rightarrow +\infty} \|x(t)\| \leq \limsup_{t \rightarrow +\infty} \|x(t)\| \leq \|x^*\|,$$

hence we have

$$\lim_{t \rightarrow +\infty} \|x(t)\| = \|x^*\|.$$

Combining this property with $x(t) \rightharpoonup x^*$ as $t \rightarrow +\infty$, we obtain the strong convergence, that is

$$\lim_{t \rightarrow +\infty} x(t) = x^*.$$

III. We suppose that for every $T \geq t_0$ there exists $t \geq T$ such that $\|x^*\| > \|x(t)\|$ and also there exists $s \geq T$ such that $\|x^*\| \leq \|x(s)\|$. From the continuity of x , we deduce that there exists a sequence $(t_n)_{n \in \mathbb{N}} \subseteq [t_0, +\infty)$ such that $t_n \rightarrow +\infty$ as $n \rightarrow +\infty$ and, for all $n \in \mathbb{N}$ we have

$$\|x(t_n)\| = \|x^*\|.$$

Consider $\bar{x} \in \mathcal{H}$ a weak sequential cluster point of $(x(t_n))_{n \in \mathbb{N}}$. We deduce as in case **II** that $\bar{x} = x^*$. Hence, x^* is the only weak sequential cluster point of $x(t_n)$ and consequently the sequence $x(t_n)$ converges weakly to x^* . Obviously $\|x(t_n)\| \rightarrow \|x^*\|$ as $n \rightarrow +\infty$. So, it follows that $x(t_n) \rightarrow x^*$, $n \rightarrow +\infty$, that is $\|x(t_n) - x^*\| \rightarrow 0$ as $n \rightarrow +\infty$. This leads to $\liminf_{t \rightarrow +\infty} \|x(t) - x^*\| = 0$.

4 Strong convergence-General case

We are going to analyze via Lyapunov analysis the strong convergence properties as $t \rightarrow +\infty$ of the solution trajectories of the inertial dynamic (TRIGS) that we recall below

$$\ddot{x}(t) + \delta \sqrt{\epsilon(t)} \dot{x}(t) + \nabla f(x(t)) + \epsilon(t)x(t) = 0.$$

Theorem 10 *Let consider the dynamic system (TRIGS) where we assume that $\epsilon(\cdot)$ satisfies the condition $(CD)_K$ for some $K > 0$, $\int_{t_0}^{+\infty} \epsilon^{\frac{3}{2}}(t) dt < +\infty$ and $\lim_{t \rightarrow +\infty} \frac{1}{\sqrt{\epsilon(t)} \exp\left(\int_{t_0}^t (\delta - K) \sqrt{\epsilon(s)} ds\right)} = 0$.*

Then, for any global solution trajectory $x : [t_0, +\infty[\rightarrow \mathcal{H}$ of (TRIGS),

$$\liminf_{t \rightarrow +\infty} \|x(t) - x^*\| = 0,$$

where x^ is the element of minimal norm of $\operatorname{argmin} f$, that is $x^* = \operatorname{proj}_{\operatorname{argmin} f} 0$.*

Further, if there exists $T \geq t_0$, such that the trajectory $\{x(t) : t \geq T\}$ stays either in the open ball $B(0, \|x^\|)$ or in its complement, then $x(t)$ converges strongly to x^* as $t \rightarrow +\infty$.*

Proof The proof is parallel to that of Theorem 9. We analyze the behavior of the trajectory $x(\cdot)$ depending on its position with respect to the ball $B(0, \|x^*\|)$.

I. Assume that $\|x(t)\| \geq \|x^*\|$ for all $t \geq T$. Let us denote $f_t(x) = f(x) + \frac{\epsilon(t)}{2} \|x\|^2$, and consider the energy functional $E : [t_1, +\infty[\rightarrow \mathbb{R}$ defined by

$$E(t) := f_t(x(t)) - f_t(x^*) + \frac{1}{2} \|c(t)(x(t) - x^*) + \dot{x}(t)\|^2,$$

where $c(t) = K\sqrt{\epsilon(t)}$. Note that $E(t) = \mathcal{E}(t) - \frac{\epsilon(t)}{2}\|x^*\|^2$, where $\mathcal{E}(t)$ was defined in the proof of Theorem 5. Hence, reasoning as in the proof of Theorem 5, see (25) (and keeping the term containing $\|x(t)\|^2$ in the right hand side of (22)), we get for all $t \geq t_1$ that

$$\dot{E}(t) + \mu(t)E(t) \leq \left(\frac{\dot{\epsilon}(t)}{2} - c(t)\frac{\epsilon(t)}{2} + \mu(t)\frac{\epsilon(t)}{2} \right) (\|x(t)\|^2 - \|x^*\|^2), \quad (70)$$

where $\mu(t) = -\frac{\dot{\epsilon}(t)}{2\epsilon(t)} + (\delta - K)\sqrt{\epsilon(t)}$. An elementary computation gives $\frac{\dot{\epsilon}(t)}{2} - c(t)\frac{\epsilon(t)}{2} + \mu(t)\frac{\epsilon(t)}{2} \leq 0$, because of $\epsilon(\cdot)$ decreasing and $K \geq \frac{\delta}{2}$. Since $\|x(t)\| \geq \|x^*\|$ for all $t \geq T$, (70) yields

$$\dot{E}(t) + \mu(t)E(t) \leq 0, \text{ for all } t \geq T_1 = \max\{T, t_1\}. \quad (71)$$

Set

$$\mathfrak{M}(t) = \exp\left(\int_{T_1}^t \mu(s)ds\right) = \exp\left(\int_{T_1}^t -\frac{\dot{\epsilon}(s)}{2\epsilon(s)} + (\delta - K)\sqrt{\epsilon(s)}ds\right).$$

Therefore, we have with $C = \sqrt{\epsilon(T_1)}$

$$\mathfrak{M}(t) = C \frac{1}{\sqrt{\epsilon(t)}} \exp\left(\int_{T_1}^t (\delta - K)\sqrt{\epsilon(s)}ds\right).$$

Multiplying (71) with $\mathfrak{M}(t)$ and integrating on an interval $[T_1, t]$, we get for all $t \geq T_1$ that

$$\mathfrak{M}(t)E(t) \leq \mathfrak{M}(T_1)E(T_1) = C'.$$

Consequently, there exists $C'_1 > 0$ such that for all $t \geq T_1$ one has

$$E(t) \leq \frac{C'_1 \sqrt{\epsilon(t)}}{\exp\left(\int_{T_1}^t (\delta - K)\sqrt{\epsilon(s)}ds\right)}.$$

Further, $f_t(x(t)) - f_t(x^*) \leq E(t)$, for all $t \geq t_1$. Therefore,

$$f_t(x(t)) - f_t(x^*) \leq \frac{C'_1 \sqrt{\epsilon(t)}}{\exp\left(\int_{T_1}^t (\delta - K)\sqrt{\epsilon(s)}ds\right)}, \text{ for all } t \geq T_1. \quad (72)$$

For fixed t let us denote $x_{\epsilon(t)} = \operatorname{argmin} f_t(x)$. Obviously $\|x_{\epsilon(t)}\| \leq \|x^*\|$.

Using the gradient inequality for the strongly convex function f_t we have

$$f_t(x) - f_t(x_{\epsilon(t)}) \geq \frac{\epsilon(t)}{2} \|x - x_{\epsilon(t)}\|^2 \text{ for all } x \in \mathcal{H} \text{ and } t \geq t_0.$$

On the other hand

$$f_t(x_{\epsilon(t)}) - f_t(x^*) = f(x_{\epsilon(t)}) - f^* + \frac{\epsilon(t)}{2} (\|x_{\epsilon(t)}\|^2 - \|x^*\|^2) \geq \frac{\epsilon(t)}{2} (\|x_{\epsilon(t)}\|^2 - \|x^*\|^2).$$

Now, by adding the last two inequalities we get

$$f_t(x) - f_t(x^*) \geq \frac{\epsilon(t)}{2} (\|x - x_{\epsilon(t)}\|^2 + \|x_{\epsilon(t)}\|^2 - \|x^*\|^2) \text{ for all } x \in \mathcal{H} \text{ and } t \geq t_0. \quad (73)$$

Hence, (72) and (73) lead to

$$\|x(t) - x_{\epsilon(t)}\|^2 + \|x_{\epsilon(t)}\|^2 - \|x^*\|^2 \leq \frac{C'_2}{\sqrt{\epsilon(t)} \exp\left(\int_{T_1}^t (\delta - K)\sqrt{\epsilon(s)}ds\right)}, \text{ for all } t \geq T_1. \quad (74)$$

Now, by taking the limit as $t \rightarrow +\infty$, and using that $x_{\epsilon(t)} \rightarrow x^*$ as $t \rightarrow +\infty$ and the assumption in the hypotheses of the theorem we get $\lim_{t \rightarrow +\infty} \|x(t) - x_{\epsilon(t)}\| \leq 0$, and hence $\lim_{t \rightarrow +\infty} x(t) = x^*$.

II. Assume now, that $\|x(t)\| < \|x^*\|$ for all $t \geq T$. By Corollary 1 we get that $f(x(t)) \rightarrow \min f$ as $t \rightarrow +\infty$. Now, we take $\bar{x} \in \mathcal{H}$ a weak sequential cluster point of the trajectory x , which exists since the trajectory is bounded. This means that there exists a sequence $(t_n)_{n \in \mathbb{N}} \subseteq [T, +\infty)$ such that $t_n \rightarrow +\infty$ and $x(t_n)$ converges weakly to \bar{x} as $n \rightarrow +\infty$. We know that f is weakly lower semicontinuous, so one has

$$f(\bar{x}) \leq \liminf_{n \rightarrow +\infty} f(x(t_n)) = \min f,$$

hence $\bar{x} \in \operatorname{argmin} f$. Now, since the norm is weakly lower semicontinuous one has that

$$\|\bar{x}\| \leq \liminf_{n \rightarrow +\infty} \|x(t_n)\| \leq \|x^*\|$$

which, from the definition of x^* , implies that $\bar{x} = x^*$. This shows that the trajectory $x(\cdot)$ converges weakly to x^* . So

$$\|x^*\| \leq \liminf_{t \rightarrow +\infty} \|x(t)\| \leq \limsup_{t \rightarrow +\infty} \|x(t)\| \leq \|x^*\|,$$

hence we have

$$\lim_{t \rightarrow +\infty} \|x(t)\| = \|x^*\|.$$

From the previous relation and the fact that $x(t) \rightarrow x^*$ as $t \rightarrow +\infty$, we obtain the strong convergence, that is

$$\lim_{t \rightarrow +\infty} x(t) = x^*.$$

III. We suppose that for every $T \geq t_0$ there exists $t \geq T$ such that $\|x^*\| > \|x(t)\|$ and also there exists $s \geq T$ such that $\|x^*\| \leq \|x(s)\|$. From the continuity of x , we deduce that there exists a sequence $(t_n)_{n \in \mathbb{N}} \subseteq [t_0, +\infty)$ such that $t_n \rightarrow +\infty$ as $n \rightarrow +\infty$ and, for all $n \in \mathbb{N}$ we have

$$\|x(t_n)\| = \|x^*\|.$$

Consider $\bar{x} \in \mathcal{H}$ a weak sequential cluster point of $(x(t_n))_{n \in \mathbb{N}}$. We deduce as at case **II** that $\bar{x} = x^*$. Hence, x^* is the only weak sequential cluster point of $x(t_n)$ and consequently the sequence $x(t_n)$ converges weakly to x^* .

Obviously $\|x(t_n)\| \rightarrow \|x^*\|$ as $n \rightarrow +\infty$. So, it follows that $x(t_n) \rightarrow x^*$, $n \rightarrow +\infty$, that is $\|x(t_n) - x^*\| \rightarrow 0$ as $n \rightarrow +\infty$. This leads to $\liminf_{t \rightarrow +\infty} \|x(t) - x^*\| = 0$.

4.1 The case $\epsilon(t)$ is of order $1/t^r$, $\frac{2}{3} < r < 2$

Take $\epsilon(t) = 1/t^r$, $\frac{2}{3} < r < 2$. Then, $\int_{t_0}^{+\infty} \epsilon^{\frac{3}{2}}(t) dt = \int_{t_0}^{+\infty} \frac{1}{t^{\frac{3}{2}r}} dt < +\infty$, $\left(\frac{1}{\sqrt{\epsilon(t)}}\right)' = \frac{r}{2} t^{\frac{r}{2}-1}$ and

$$\lim_{t \rightarrow +\infty} \frac{1}{\sqrt{\epsilon(t)} \exp\left(\int_{t_0}^t (\delta - K) \sqrt{\epsilon(s)} ds\right)} = \lim_{t \rightarrow +\infty} \frac{Ct^{\frac{r}{2}}}{\exp\left(\frac{2(\delta-K)}{2-r} t^{1-\frac{r}{2}}\right)} = 0.$$

Therefore, Theorem 10 can be applied. Let us summarize these results in the following statement.

Theorem 11 Take $\epsilon(t) = 1/t^r$, $\frac{2}{3} < r < 2$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a global solution trajectory of

$$\ddot{x}(t) + \frac{\delta}{t^{\frac{r}{2}}} \dot{x}(t) + \nabla f(x(t)) + \frac{1}{t^r} x(t) = 0.$$

Then, $\liminf_{t \rightarrow +\infty} \|x(t) - x^*\| = 0$.

Further, if there exists $T \geq t_0$, such that the trajectory $\{x(t) : t \geq T\}$ stays either in the open ball $B(0, \|x^*\|)$ or in its complement, then $x(t)$ converges strongly to x^* as $t \rightarrow +\infty$.

5 Fast inertial algorithms with Tikhonov regularization

On the basis of the convergence properties of continuous dynamic (TRIGS), one would expect to obtain similar results for the algorithms resulting from its temporal discretization. To illustrate this, we will do a detailed study of the associated proximal algorithms, obtained by implicit discretization. A full study of the associated first-order algorithms would be beyond the scope of this article, and will be the subject of further study. So, for $k \geq 1$, consider the discrete dynamic

$$(x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha}{k}(x_k - x_{k-1}) + \nabla f(x_{k+1}) + \frac{c}{k^2}\xi_k = 0, \quad (75)$$

with time step size equal to one. We take $\xi_k = x_k$, which gives

$$\text{(IPATRE)} \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_f\left(y_k - \frac{c}{k^2}x_k\right), \end{cases}$$

where (IPATRE) stands for Inertial Proximal Algorithm with Tikhonov REgularization. According to (75) we have

$$x_{k+1} = \alpha_k(x_k - x_{k-1}) - \nabla f(x_{k+1}) + \left(1 - \frac{c}{k^2}\right)x_k. \quad (76)$$

5.1 Convergence of values

We have the following result.

Theorem 12 *Let (x_k) be a sequence generated by (IPATRE). Assume that $\alpha > 3$. Then for all $s \in [\frac{1}{2}, 1[$ the following hold:*

- (i) $f(x_k) - \min_{\mathcal{H}} f = o(k^{-2s})$, $\|x_k - x_{k-1}\| = o(k^{-s})$ and $\|\nabla f(x_k)\| = o(k^{-s})$ as $k \rightarrow +\infty$.
- (ii) $\sum_{k=1}^{+\infty} k^{2s-1}(f(x_k) - \min_{\mathcal{H}} f) < +\infty$, $\sum_{k=1}^{+\infty} k^{2s-1}\|x_k - x_{k-1}\|^2 < +\infty$, $\sum_{k=1}^{+\infty} k^{2s}\|\nabla f(x_k)\|^2 < +\infty$.

Proof Given $x^* \in \text{argmin } f$, set $f^* = f(x^*) = \min_{\mathcal{H}} f$. For $k \geq 2$, consider the discrete energy

$$E_k := \|a_{k-1}(x_{k-1} - x^*) + b_{k-1}(x_k - x_{k-1} + \nabla f(x_k))\|^2 + d_{k-1}\|x_{k-1}\|^2, \quad (77)$$

where $a_k = ak^{r-1}$, $2 < a < \alpha - 1$ and $b_k = k^r$, $r \in]0, 1]$. The sequence (d_k) will be defined later. Set shortly $c_k := \frac{c}{k^2}$. Let us develop E_k .

$$\begin{aligned} E_k &= a_{k-1}^2\|x_{k-1} - x^*\|^2 + b_{k-1}^2\|x_k - x_{k-1}\|^2 + b_{k-1}^2\|\nabla f(x_k)\|^2 + 2a_{k-1}b_{k-1}\langle x_k - x_{k-1}, x_{k-1} - x^* \rangle \\ &\quad + 2a_{k-1}b_{k-1}\langle \nabla f(x_k), x_{k-1} - x^* \rangle + 2b_{k-1}^2\langle \nabla f(x_k), x_k - x_{k-1} \rangle + d_{k-1}\|x_{k-1}\|^2. \end{aligned} \quad (78)$$

Further

$$\begin{aligned} 2a_{k-1}b_{k-1}\langle x_k - x_{k-1}, x_{k-1} - x^* \rangle &= a_{k-1}b_{k-1}(\|x_k - x^*\|^2 - \|x_k - x_{k-1}\|^2 - \|x_{k-1} - x^*\|^2) \\ 2a_{k-1}b_{k-1}\langle \nabla f(x_k), x_{k-1} - x^* \rangle &= 2a_{k-1}b_{k-1}\langle \nabla f(x_k), x_k - x^* \rangle - 2a_{k-1}b_{k-1}\langle \nabla f(x_k), x_k - x_{k-1} \rangle. \end{aligned}$$

Consequently, (78) becomes

$$\begin{aligned} E_k &= a_{k-1}b_{k-1}\|x_k - x^*\|^2 + (a_{k-1}^2 - a_{k-1}b_{k-1})\|x_{k-1} - x^*\|^2 + (b_{k-1}^2 - a_{k-1}b_{k-1})\|x_k - x_{k-1}\|^2 \\ &\quad + b_{k-1}^2\|\nabla f(x_k)\|^2 + 2a_{k-1}b_{k-1}\langle \nabla f(x_k), x_k - x^* \rangle + (2b_{k-1}^2 - 2a_{k-1}b_{k-1})\langle \nabla f(x_k), x_k - x_{k-1} \rangle \\ &\quad + d_{k-1}\|x_{k-1}\|^2. \end{aligned} \quad (79)$$

Let us proceed similarly with E_{k+1} . Let us first observe that from (77) we have

$$E_{k+1} = \|a_k(x_k - x^*) + b_k(\alpha_k(x_k - x_{k-1}) - c_k x_k)\|^2 + d_k \|x_k\|^2.$$

Therefore, after development we get

$$\begin{aligned} E_{k+1} &= a_k^2 \|x_k - x^*\|^2 + \alpha_k^2 b_k^2 \|x_k - x_{k-1}\|^2 + b_k^2 c_k^2 \|x_k\|^2 + 2\alpha_k a_k b_k \langle x_k - x_{k-1}, x_k - x^* \rangle \\ &\quad - 2\alpha_k b_k^2 c_k \langle x_k - x_{k-1}, x_k \rangle - 2a_k b_k c_k \langle x_k, x_k - x^* \rangle + d_k \|x_k\|^2. \end{aligned} \quad (80)$$

Further,

$$\begin{aligned} 2\alpha_k a_k b_k \langle x_k - x_{k-1}, x_k - x^* \rangle &= -\alpha_k a_k b_k (\|x_{k-1} - x^*\|^2 - \|x_k - x_{k-1}\|^2 - \|x_k - x^*\|^2) \\ -2\alpha_k b_k^2 c_k \langle x_k - x_{k-1}, x_k \rangle &= \alpha_k b_k^2 c_k (\|x_{k-1}\|^2 - \|x_k - x_{k-1}\|^2 - \|x_k\|^2) \\ -2a_k b_k c_k \langle x_k, x_k - x^* \rangle &= a_k b_k c_k (\|x^*\|^2 - \|x_k - x^*\|^2 - \|x_k\|^2). \end{aligned}$$

Therefore, (80) yields

$$\begin{aligned} E_{k+1} &= (a_k^2 + \alpha_k a_k b_k - a_k b_k c_k) \|x_k - x^*\|^2 - \alpha_k a_k b_k \|x_{k-1} - x^*\|^2 \\ &\quad + (\alpha_k^2 b_k^2 + \alpha_k a_k b_k - \alpha_k b_k^2 c_k) \|x_k - x_{k-1}\|^2 + (b_k^2 c_k^2 + d_k - \alpha_k b_k^2 c_k - a_k b_k c_k) \|x_k\|^2 \\ &\quad + \alpha_k b_k^2 c_k \|x_{k-1}\|^2 + a_k b_k c_k \|x^*\|^2. \end{aligned} \quad (81)$$

By combining (79) and (81), we obtain

$$\begin{aligned} E_{k+1} - E_k &= (a_k^2 + \alpha_k a_k b_k - a_k b_k c_k - a_{k-1} b_{k-1}) \|x_k - x^*\|^2 \\ &\quad + (-\alpha_k a_k b_k - a_{k-1}^2 + a_{k-1} b_{k-1}) \|x_{k-1} - x^*\|^2 \\ &\quad + (\alpha_k^2 b_k^2 + \alpha_k a_k b_k - \alpha_k b_k^2 c_k - b_{k-1}^2 + a_{k-1} b_{k-1}) \|x_k - x_{k-1}\|^2 \\ &\quad + (b_k^2 c_k^2 + d_k - \alpha_k b_k^2 c_k - a_k b_k c_k) \|x_k\|^2 + (\alpha_k b_k^2 c_k - d_{k-1}) \|x_{k-1}\|^2 - b_{k-1}^2 \|\nabla f(x_k)\|^2 \\ &\quad + 2a_{k-1} b_{k-1} \langle \nabla f(x_k), x^* - x_k \rangle + (2b_{k-1}^2 - 2a_{k-1} b_{k-1}) \langle \nabla f(x_k), x_{k-1} - x_k \rangle + a_k b_k c_k \|x^*\|^2. \end{aligned} \quad (82)$$

By convexity of f , we have

$$\langle \nabla f(x_k), x^* - x_k \rangle \leq f^* - f(x_k) \text{ and } \langle \nabla f(x_k), x_{k-1} - x_k \rangle \leq f(x_{k-1}) - f(x_k).$$

According to the form of (a_k) and (b_k) , there exists $k_0 \geq 2$ such that $b_k \geq a_k$ for all $k \geq k_0$. Consequently, $2b_{k-1}^2 - 2a_{k-1} b_{k-1} \geq 0$ which, according to the above convexity inequalities, gives

$$\begin{aligned} &2a_{k-1} b_{k-1} \langle \nabla f(x_k), x^* - x_k \rangle + (2b_{k-1}^2 - 2a_{k-1} b_{k-1}) \langle \nabla f(x_k), x_{k-1} - x_k \rangle \\ &\leq 2a_{k-1} b_{k-1} (f^* - f(x_k)) + (2b_{k-1}^2 - 2a_{k-1} b_{k-1}) [f(x_{k-1}) - f(x_k)] \\ &= -2a_{k-1} b_{k-1} (f(x_k) - f^*) + (2b_{k-1}^2 - 2a_{k-1} b_{k-1}) [(f(x_{k-1}) - f^*) - (f(x_k) - f^*)] \\ &= (2b_{k-1}^2 - 2a_{k-1} b_{k-1}) (f(x_{k-1}) - f^*) - 2b_{k-1}^2 (f(x_k) - f^*) \\ &= (2b_{k-1}^2 - 2a_{k-1} b_{k-1}) (f(x_{k-1}) - f^*) - \left((2b_k^2 - 2a_k b_k) + (2b_{k-1}^2 - 2b_k^2 + 2a_k b_k) \right) (f(x_k) - f^*). \end{aligned} \quad (83)$$

Set $\mu_k := 2b_k^2 - 2a_k b_k$ and observe that $\mu_k \geq 0$ for all $k \geq k_0$, and $\mu_k \sim Ck^{2r}$ (we use C as a generic positive constant). Let us also introduce $m_k := 2b_{k-1}^2 - 2b_k^2 + 2a_k b_k$, and observe that $m_k \geq 0$ for all $k \geq k_0$. Equivalently, let us show that for all $\frac{1}{2} \leq r \leq 1$ one has $b_k^2 - a_k b_k \leq b_{k-1}^2$ for all $k \geq 1$. Equivalently $k^{2r} - ak^{2r-1} - (k-1)^{2r} \leq 0$. By convexity of the function $x \mapsto x^{2r}$, the subgradient inequality gives

$$(x-1)^{2r} \geq x^{2r} - 2rx^{2r-1} \geq x^{2r} - ax^{2r-1},$$

where the second inequality comes from $2r < a$. Replacing x with k gives the claim. In addition $m_k \sim Ck^{2r-1}$. Combining (82) and (83), we obtain that for all $k \geq k_0$

$$\begin{aligned}
& E_{k+1} - E_k + \mu_k(f(x_k) - f^*) - \mu_{k-1}(f(x_{k-1}) - f^*) + m_k(f(x_k) - f^*) \\
& \leq (a_k^2 + \alpha_k a_k b_k - a_k b_k c_k - a_{k-1} b_{k-1}) \|x_k - x^*\|^2 \\
& \quad + (-\alpha_k a_k b_k - a_{k-1}^2 + a_{k-1} b_{k-1}) \|x_{k-1} - x^*\|^2 \\
& \quad + (\alpha_k^2 b_k^2 + \alpha_k a_k b_k - \alpha_k b_k^2 c_k - b_{k-1}^2 + a_{k-1} b_{k-1}) \|x_k - x_{k-1}\|^2 \\
& \quad + (b_k^2 c_k^2 + d_k - \alpha_k b_k^2 c_k - a_k b_k c_k) \|x_k\|^2 + (\alpha_k b_k^2 c_k - d_{k-1}) \|x_{k-1}\|^2 - b_{k-1}^2 \|\nabla f(x_k)\|^2 \\
& \quad + a_k b_k c_k \|x^*\|^2.
\end{aligned} \tag{84}$$

Let us now analyze the right hand side of (84).

i) Write the coefficient of $\|x_k - x^*\|^2$ so as to show a term similar to the coefficient of $\|x_{k-1} - x^*\|^2$. This will prepare the summation of these quantities. This gives

$$\begin{aligned}
& a_k^2 + \alpha_k a_k b_k - a_k b_k c_k - a_{k-1} b_{k-1} = (\alpha_{k+1} a_{k+1} b_{k+1} + a_k^2 - a_k b_k) \\
& \quad + (\alpha_k a_k b_k - a_k b_k c_k - a_{k-1} b_{k-1} - \alpha_{k+1} a_{k+1} b_{k+1} + a_k b_k).
\end{aligned} \tag{85}$$

a) By definition, $\alpha_{k+1} a_{k+1} b_{k+1} + a_k^2 - a_k b_k = a(k+1)^{2r-1} - \alpha a(k+1)^{2r-2} + a^2 k^{2r-2} - a k^{2r-1}$. Proceeding as before, let us show that $a(x+1)^{2r-1} - \alpha a(x+1)^{2r-2} + a^2 x^{2r-2} - a x^{2r-1} \leq 0$ for x large enough. By taking $\frac{1}{2} \leq r \leq 1$, by convexity of the function $x \mapsto -x^{2r-1}$, the subgradient inequality gives $(2r-1)x^{2r-2} \geq (x+1)^{2r-1} - x^{2r-1}$. Therefore,

$$a(x+1)^{2r-1} - a x^{2r-1} - \alpha a(x+1)^{2r-2} + a^2 x^{2r-2} \leq a(2r-1)x^{2r-2} - \alpha a(x+1)^{2r-2} + a^2 x^{2r-2}.$$

But $a(2r-1)x^{2r-2} + a^2 x^{2r-2} \leq \alpha a(x+1)^{2r-2}$ since $2r+a \leq \alpha+1$ and the claim follows.

Therefore, there exists $k_1 \geq k_0$ such that for all $\frac{1}{2} \leq r \leq 1$ we have

$$\alpha_{k+1} a_{k+1} b_{k+1} + a_k^2 - a_k b_k \leq 0, \text{ for all } k \geq k_1. \tag{86}$$

Set $\nu_k := -\alpha_{k+1} a_{k+1} b_{k+1} - a_k^2 + a_k b_k$. According to (86), $\nu_k \geq 0$ for all $k \geq k_1$, and $\nu_k \sim Ck^{2r-2}$.

b) Consider now the second term in the right hand side of (85):

$$\begin{aligned}
& a_k a_k b_k - a_k b_k c_k - a_{k-1} b_{k-1} - \alpha_{k+1} a_{k+1} b_{k+1} + a_k b_k \\
& = 2a k^{2r-1} - \alpha a k^{2r-2} - a c k^{2r-3} - a(k-1)^{2r-1} - a(k+1)^{2r-1} + \alpha a(k+1)^{2r-2}.
\end{aligned}$$

Let us show that for all $\frac{1}{2} \leq r \leq 1$

$$\phi(x, r) = 2a x^{2r-1} - \alpha a x^{2r-2} - a c x^{2r-3} - a(x-1)^{2r-1} - a(x+1)^{2r-1} + \alpha a(x+1)^{2r-2} \leq 0$$

for x large enough. By convexity of the function $x \mapsto x^{2r-1} - (x-1)^{2r-1}$ (one can easily verify that its second order derivative is nonnegative), the subgradient inequality gives $(x+1)^{2r-1} - 2x^{2r-1} + (x-1)^{2r-1} \geq (2r-1)(x^{2r-2} - (x-1)^{2r-2})$. Therefore

$$\begin{aligned}
\phi(x, r) & = -a[(x+1)^{2r-1} - 2x^{2r-1} + (x-1)^{2r-1}] - \alpha a x^{2r-2} - a c x^{2r-3} + \alpha a(k+1)^{2r-2} \\
& \leq -a[(2r-1)(x^{2r-2} - (x-1)^{2r-2})] - \alpha a x^{2r-2} - a c x^{2r-3} + \alpha a(k+1)^{2r-2} \\
& = a(2r-1)(x-1)^{2r-2} - a(\alpha+2r-1)x^{2r-2} - a c x^{2r-3} + \alpha a(x+1)^{2r-2}.
\end{aligned}$$

Similarly, by convexity of the function $x \mapsto (x-1)^{2r-2} - x^{2r-2}$, the subgradient inequality gives $2x^{2r-2} - (x+1)^{2r-2} - (x-1)^{2r-2} \geq (2r-2)((x-1)^{2r-3} - x^{2r-3})$. Therefore, $\alpha a(x+1)^{2r-2} - \alpha a x^{2r-2} \leq \alpha a(x^{2r-2} - (x-1)^{2r-2}) - \alpha a(2r-2)((x-1)^{2r-3} - x^{2r-3})$. Consequently,

$$\phi(x, r) \leq a(2r-1-\alpha)((x-1)^{2r-2} - x^{2r-2}) - \alpha a(2r-2)((x-1)^{2r-3} - x^{2r-3}) - a c x^{2r-3}.$$

Finally, by convexity of the function $x \mapsto x^{2r-2}$, the subgradient inequality gives $(x-1)^{2r-2} - x^{2r-2} \geq -(2r-2)x^{2r-3}$. Taking into account that $a(2r-1-\alpha) \leq 0$ we get

$$\phi(x, r) \leq -a(2r-1-2\alpha)(2r-2)x^{2r-3} - a\alpha(2r-2)(x-1)^{2r-3} - acx^{2r-3}.$$

Since $\frac{2\alpha+1-2r}{\alpha} > 1$ we obtain that $\phi(x, r) \leq 0$ for $x > 1$.

Consequently, there exists $k_2 \geq k_1$ such that for all $\frac{1}{2} \leq r \leq 1$

$$\alpha_k a_k b_k - a_k b_k c_k - a_{k-1} b_{k-1} - \alpha_{k+1} a_{k+1} b_{k+1} + a_k b_k \leq 0, \text{ for all } k \geq k_2. \quad (87)$$

Set $n_k := -\alpha_k a_k b_k + a_k b_k c_k + a_{k-1} b_{k-1} + \alpha_{k+1} a_{k+1} b_{k+1} - a_k b_k$. So $n_k \geq 0$ for all $k \geq k_2$ and $n_k \sim Ck^{2r-3}$.

ii) Let us now examine the coefficient of $\|x_k - x_{k-1}\|^2$. By definition we have

$$\begin{aligned} & \alpha_k^2 b_k^2 + \alpha_k a_k b_k - \alpha_k b_k^2 c_k - b_{k-1}^2 + a_{k-1} b_{k-1} \\ & = k^{2r} - (k-1)^{2r} + (-2\alpha + a)k^{2r-1} + a(k-1)^{2r-1} + (\alpha^2 - \alpha a - c)k^{2r-2} + \alpha c k^{2r-3}. \end{aligned}$$

Let us show that for all $\frac{1}{2} \leq r \leq 1$

$$\phi(x, r) = x^{2r} - (x-1)^{2r} + (-2\alpha + a)x^{2r-1} + a(x-1)^{2r-1} + (\alpha^2 - \alpha a - c)x^{2r-2} + acx^{2r-3} \leq 0,$$

if x is large enough. By convexity of the function $x \mapsto x^{2r} - ax^{2r-1}$, the subgradient inequality gives $((x-1)^{2r} - a(x-1)^{2r-1}) - (x^{2r} - ax^{2r-1}) \geq -(2rx^{2r-1} - a(2r-1)x^{2r-2})$. Therefore, taking into account that $r - \alpha + a \leq 1 - \alpha + a \leq 0$, we obtain

$$\phi(x, r) \leq 2(r - \alpha + a)x^{2r-1} - a(2r-1)x^{2r-2} + (\alpha^2 - \alpha a - c)x^{2r-2} + acx^{2r-3} \leq 0,$$

for x large enough. Consequently, there exist $k_3 \geq k_2$ such that for all $\frac{1}{2} \leq r \leq 1$

$$\alpha_k^2 b_k^2 + \alpha_k a_k b_k - \alpha_k b_k^2 c_k - b_{k-1}^2 + a_{k-1} b_{k-1} \leq 0, \text{ for all } k \geq k_3. \quad (88)$$

Set $\eta_k := -\alpha_k^2 b_k^2 - \alpha_k a_k b_k + \alpha_k b_k^2 c_k + b_{k-1}^2 - a_{k-1} b_{k-1}$. So $\eta_k \geq 0$ for all $k \geq k_3$ and $\eta_k \sim Ck^{2r-1}$.

iii) The coefficient of $\|x_{k-1}\|^2$ is $\alpha_k b_k^2 c_k - d_{k-1}$. We proceed in a similar way as in *i*), and write the coefficient of $\|x_k\|^2$ as

$$b_k^2 c_k^2 + d_k - \alpha_k b_k^2 c_k - a_k b_k c_k = (-\alpha_{k+1} b_{k+1}^2 c_{k+1} + d_k) + (b_k^2 c_k^2 + \alpha_{k+1} b_{k+1}^2 c_{k+1} - \alpha_k b_k^2 c_k - a_k b_k c_k).$$

We have

$$\begin{aligned} & b_k^2 c_k^2 + \alpha_{k+1} b_{k+1}^2 c_{k+1} - \alpha_k b_k^2 c_k - a_k b_k c_k = c^2 k^{2r-4} + c(k+1)^{2r-2} - \alpha c(k+1)^{2r-3} \\ & \quad - ck^{2r-2} + \alpha ck^{2r-3} - ack^{2r-3}. \end{aligned}$$

Let us show that for all $\frac{1}{2} \leq r \leq 1$

$$\phi(x, r) = c(x+1)^{2r-2} - \alpha c(x+1)^{2r-3} - cx^{2r-2} + \alpha cx^{2r-3} - acx^{2r-3} + c^2 x^{2r-4} \leq 0$$

for x large enough. Since for x large enough, the function $x \mapsto x^{2r-2} - \alpha x^{2r-3}$ is convex, the subgradient inequality gives

$$(x^{2r-2} - \alpha x^{2r-3}) - ((x+1)^{2r-2} - \alpha(x+1)^{2r-3}) \geq -((2r-2)(x+1)^{2r-3} - \alpha(2r-3)(x+1)^{2r-4}).$$

Therefore, by taking into account that $r \leq 1$, we obtain

$$\phi(x, r) \leq (2r-2)c(x+1)^{2r-3} - \alpha(2r-3)c(x+1)^{2r-4} - acx^{2r-3} + c^2 x^{2r-4} \leq 0$$

for x large enough. Consequently, there exists $k_4 \geq k_3$ such that for all $\frac{1}{2} \leq r \leq 1$ we have

$$b_k^2 c_k^2 + \alpha_{k+1} b_{k+1}^2 c_{k+1} - \alpha_k b_k^2 c_k - a_k b_k c_k \leq 0 \text{ for all } k \geq k_4. \quad (89)$$

Let us denote $\sigma_k := \alpha_{k+1} b_{k+1}^2 c_{k+1} - d_k$ and $s_k := -b_k^2 c_k^2 - \alpha_{k+1} b_{k+1}^2 c_{k+1} + \alpha_k b_k^2 c_k + a_k b_k c_k$ and observe that $s_k \geq 0$ for all $k \geq k_4$ and $s_k \sim Ck^{2r-3}$.

Combining (84), (86), (87), (88) and (89) we obtain that for all $k \geq k_4$ and $r \in [\frac{1}{2}, 1]$ it holds

$$\begin{aligned} & E_{k+1} - E_k + \mu_k(f(x_k) - f^*) - \mu_{k-1}(f(x_{k-1}) - f^*) + m_k(f(x_k) - f^*) \\ & + \nu_k \|x_k - x^*\|^2 - \nu_{k-1} \|x_{k-1} - x^*\|^2 + n_k \|x_k - x^*\|^2 \\ & + \sigma_k \|x_k\|^2 - \sigma_{k-1} \|x_{k-1}\|^2 + s_k \|x_k\|^2 \\ & + \eta_k \|x_k - x_{k-1}\|^2 + b_{k-1}^2 \|\nabla f(x_k)\|^2 \leq a_k b_k c_k \|x^*\|^2. \end{aligned} \quad (90)$$

Finally, take $d_{k-1} = \frac{1}{2} \alpha_k b_k^2 c_k$. Then, $\sigma_k = \frac{1}{2} \alpha_{k+1} b_{k+1}^2 c_{k+1} \sim Ck^{2r-2}$, $\sigma_k \geq 0$ for all $k \geq k_5 = \max(\alpha - 1, k_4)$. Further, $\mu_k, m_k, \nu_k, n_k, s_k$ and η_k are nonnegative for all $k \geq k_5$ and $r \in [\frac{1}{2}, 1]$.

Assume now that $\frac{1}{2} \leq r < 1$. According to $\sum_{k \geq k_5} a_k b_k c_k \|x^*\|^2 = ac \|x^*\|^2 \sum_{k \geq k_5} k^{2r-3} = C < +\infty$, by summing up (90) from $k = k_5$ to $k = n > k_5$, we obtain that there exists $C_1 > 0$ such that

$$\begin{aligned} & E_{n+1} \leq C_1, \\ & \mu_n(f(x_n) - f^*) \leq C_1, \text{ hence } f(x_n) - f^* = \mathcal{O}(n^{-2r}), \\ & \sum_{k \geq k_5} m_k(f(x_k) - f^*) \leq C_1, \text{ hence } \sum_{k \geq 1} k^{2r-1}(f(x_k) - f^*) < +\infty, \\ & \nu_k \|x_k - x^*\|^2 \leq C_1, \text{ hence } \|x_n - x^*\| = \mathcal{O}(n^{1-r}), \\ & \sum_{k \geq k_5} n_k \|x_k - x^*\|^2 \leq C_1, \text{ hence } \sum_{k \geq 1} k^{2r-3} \|x_k - x^*\|^2 < +\infty, \\ & \sigma_k \|x_k\|^2 \leq C_1, \text{ hence } \|x_n\| = \mathcal{O}(n^{1-r}), \\ & \sum_{k \geq k_5} s_k \|x_k\|^2 \leq C_1, \text{ hence } \sum_{k \geq 1} k^{2r-3} \|x_k\|^2 < +\infty, \\ & \sum_{k \geq k_5} \eta_k \|x_k - x_{k-1}\|^2 \leq C_1, \text{ hence } \sum_{k \geq 1} k^{2r-1} \|x_k - x_{k-1}\|^2 < +\infty \\ & \sum_{k \geq k_5} b_{k-1}^2 \|\nabla f(x_k)\|^2 \leq C_1, \text{ hence } \sum_{k \geq 1} k^{2r} \|\nabla f(x_k)\|^2 < +\infty. \end{aligned}$$

Since $\sum_{k \geq 1} k^{2r} \|\nabla f(x_k)\|^2 < +\infty$, we have $\|\nabla f(x_n)\| = o(n^{-r})$. Combining this property with $E_{n+1} \leq C_1$ yields $\sup_{n \geq 1} \|an^{r-1}(x_n - x^*) + n^r(x_{n+1} - x_n)\| + \frac{c}{2} (1 - \frac{\alpha}{n}) n^{2r-2} \|x_{n-1}\|^2 < +\infty$.

Let us show now, that $f(x_n) - f^* = o(n^{-2r})$ and $\|x_n - x_{n-1}\| = o(n^{-r})$. From (90) we get

$$\sum_{k \geq 1} [(E_{k+1} + \mu_k(f(x_k) - f^*) + \nu_k \|x_k - x^*\|^2) - (E_k + \mu_{k-1}(f(x_{k-1}) - f^*) + \nu_{k-1} \|x_{k-1} - x^*\|^2)]_+ < +\infty.$$

Therefore, the following limit exists

$$\lim_{k \rightarrow +\infty} (\|ak^{r-1}(x_k - x^*) + k^r(x_{k+1} - x_k)\|^2 + d_k \|x_k\|^2 + \mu_k(f(x_k) - f^*) + \nu_k \|x_k - x^*\|^2).$$

Note that $d_k \sim Ck^{2r-2}$, $\mu_k \sim Ck^{2r}$ and $\nu_k \sim Ck^{2r-2}$. Further, we have $\sum_{k \geq 1} k^{2r-3} \|x_k - x^*\|^2 < +\infty$, $\sum_{k \geq 1} k^{2r-1} \|x_k - x_{k-1}\|^2 < +\infty$, $\sum_{k \geq 1} k^{2r-1} (f(x_k) - f^*) < +\infty$ and $\sum_{k \geq 1} k^{2r-3} \|x_k\|^2 < +\infty$, hence

$$\sum_{k \geq 1} \frac{1}{k} (\|ak^{r-1}(x_k - x^*) + k^r(x_{k+1} - x_k)\|^2 + d_k \|x_k\|^2 + \mu_k(f(x_k) - f^*) + \nu_k \|x_k - x^*\|^2) < +\infty.$$

Since $\sum_{k \geq 1} \frac{1}{k} = +\infty$ we get

$$\lim_{k \rightarrow +\infty} (\|ak^{r-1}(x_k - x^*) + k^r(x_{k+1} - x_k)\|^2 + d_k \|x_k\|^2 + \mu_k(f(x_k) - f^*) + \nu_k \|x_k - x^*\|^2) = 0$$

and the claim follows.

Remark 4 The convergence rate of the values is $f(x_k) - \min_{\mathcal{H}} f = o(k^{-2s})$ for any $0 < s < 1$. Practically it is as good as the rate $f(x(t)) - \min_{\mathcal{H}} f = O(\frac{1}{t^2})$ obtained for the continuous dynamic.

5.2 Strong convergence to the minimum norm solution

Theorem 13 *Take $\alpha > 3$. Let (x_k) be a sequence generated by (IPATRE). Let x^* be the minimum norm element of $\text{argmin } f$. Then, $\liminf_{k \rightarrow +\infty} \|x_k - x^*\| = 0$. Further, (x_k) converges strongly to x^* whenever (x_k) is in the interior of the ball $B(0, \|x^*\|)$ for k large enough, or (x_k) is in the complement of the ball $B(0, \|x^*\|)$ for k large enough.*

Proof Case I. Assume that there exists $k_0 \in \mathbb{N}$ such that $\|x_k\| \geq \|x^*\|$ for all $k \geq k_0$. Set $c_k = \frac{c}{k^2}$, and define $f_{c_k}(x) := f(x) + \frac{c}{2k^2} \|x\|^2$. Consider the energy function defined in (77) with $r = 1$, that is $a_k = a$ and $b_k = k^2$, where we assume that $\max(2, \alpha - 2) < a < \alpha - 1$. Then,

$$E_k = \|a(x_{k-1} - x^*) + (k-1)^2(x_k - x_{k-1} + \nabla f(x_k))\|^2 + d_{k-1} \|x_{k-1}\|^2,$$

where the sequence (d_k) will be defined later. Next, we introduce another energy functional

$$\mathcal{E}_k = \frac{1}{2} c_{k-1} (\|x_{k-1}\|^2 - \|x^*\|^2) + \|a(x_{k-1} - x^*) + (k-1)^2(x_k - x_{k-1} + \nabla f(x_k))\|^2 + d_{k-1} \|x_{k-1}\|^2. \quad (91)$$

Note that $\mathcal{E}_k = \frac{1}{2} c_{k-1} (\|x_{k-1}\|^2 - \|x^*\|^2) + E_k$. Then,

$$\mathcal{E}_{k+1} - \mathcal{E}_k = \frac{1}{2} c_k (\|x_k\|^2 - \|x^*\|^2) - \frac{1}{2} c_{k-1} (\|x_{k-1}\|^2 - \|x^*\|^2) + E_{k+1} - E_k. \quad (92)$$

According to (90), there exists $k_1 \geq k_0$ such that for all $k \geq k_1$

$$\begin{aligned} \mathcal{E}_{k+1} - \mathcal{E}_k + \mu_k(f(x_k) - f^*) - \mu_{k-1}(f(x_{k-1}) - f^*) + m_k(f(x_k) - f^*) \\ + \nu_k \|x_k - x^*\|^2 - \nu_{k-1} \|x_{k-1} - x^*\|^2 + n_k \|x_k - x^*\|^2 \\ + \eta_k \|x_k - x_{k-1}\|^2 + b_{k-1}^2 \|\nabla f(x_k)\|^2 \leq -\sigma_k \|x_k\|^2 + \sigma_{k-1} \|x_{k-1}\|^2 - s_k \|x_k\|^2 \\ + \frac{1}{2} c_k (\|x_k\|^2 - \|x^*\|^2) - \frac{1}{2} c_{k-1} (\|x_{k-1}\|^2 - \|x^*\|^2) + a_k b_k c_k \|x^*\|^2. \end{aligned} \quad (93)$$

Adding $\frac{1}{2}(\mu_k + m_k)c_k(\|x_k\|^2 - \|x^*\|^2) - \frac{1}{2}\mu_{k-1}c_{k-1}(\|x_{k-1}\|^2 - \|x^*\|^2)$ to both side of (93) we get

$$\begin{aligned} \mathcal{E}_{k+1} - \mathcal{E}_k + \mu_k(f_{c_k}(x_k) - f_{c_k}(x^*)) - \mu_{k-1}(f_{c_{k-1}}(x_{k-1}) - f_{c_{k-1}}(x^*)) + m_k(f_{c_k}(x_k) - f_{c_k}(x^*)) \\ + \nu_k \|x_k - x^*\|^2 - \nu_{k-1} \|x_{k-1} - x^*\|^2 + n_k \|x_k - x^*\|^2 + \eta_k \|x_k - x_{k-1}\|^2 + b_{k-1}^2 \|\nabla f(x_k)\|^2 \\ \leq -\sigma_k \|x_k\|^2 + \sigma_{k-1} \|x_{k-1}\|^2 - s_k \|x_k\|^2 \\ + \frac{1}{2}(\mu_k + m_k + 1)c_k(\|x_k\|^2 - \|x^*\|^2) - \frac{1}{2}(\mu_{k-1} + 1)c_{k-1}(\|x_{k-1}\|^2 - \|x^*\|^2) + a_k b_k c_k \|x^*\|^2. \end{aligned} \quad (94)$$

The right hand side of (94) can be written as

$$\begin{aligned} \left(\frac{1}{2}(\mu_k + m_k + 1)c_k - \sigma_k - s_k \right) (\|x_k\|^2 - \|x^*\|^2) \\ + \left(-\frac{1}{2}(\mu_{k-1} + 1)c_{k-1} + \sigma_{k-1} \right) (\|x_{k-1}\|^2 - \|x^*\|^2) + (a_k b_k c_k - \sigma_k - s_k + \sigma_{k-1}) \|x^*\|^2. \end{aligned}$$

In this case we have $\mu_k = 2b_k^2 - 2a_k b_k = 2k^2 - 2ak$ and $m_k = 2b_{k-1}^2 - 2b_k^2 + 2a_k b_k = 2(a-2)k + 2$. Further, $\sigma_k = \alpha_{k+1} b_{k+1}^2 c_{k+1} - d_k = c - \frac{\alpha c}{k+1} - d_k$ and $s_k = b_k^2 c_k^2 - \alpha_{k+1} b_{k+1}^2 c_{k+1} + \alpha_k b_k^2 c_k + a_k b_k c_k = \frac{\alpha c}{k+1} + \frac{c(a-\alpha)}{k} + \frac{c^2}{k^2}$. Now, take $d_k = \frac{(a+2-\alpha)c}{2k} \geq 0$ and an easy computation gives that there exists $k_2 \geq k_1$ such that for all $k \geq k_2$ one has

$$\begin{aligned} \frac{1}{2}(\mu_k + m_k + 1)c_k - \sigma_k - s_k &= -\frac{(a+2-\alpha)c}{2k} + \frac{2c^2 - 3c}{2k^2} \leq 0, \\ -\frac{1}{2}(\mu_{k-1} + 1)c_{k-1} + \sigma_{k-1} &= \frac{c(a-2-\alpha)}{2(k-1)} + \frac{\alpha c}{k(k-1)} - \frac{c}{2(k-1)^2} \leq 0 \\ a_k b_k c_k - \sigma_k - s_k + \sigma_{k-1} &= \frac{(a+2-\alpha)c}{2k} - \frac{(a+2-\alpha)c}{2(k-1)} - \frac{c^2}{k^2} \leq 0. \end{aligned}$$

Now, since by assumption $\|x_k\| \geq \|x^*\|$ for $k \geq k_0$, we get that the right hand side of (94) is nonpositive for all $k \geq k_2$. Hence, for all $k \geq k_2$ we have

$$\begin{aligned} \mathcal{E}_{k+1} - \mathcal{E}_k + \mu_k(f_{c_k}(x_k) - f_{c_k}(x^*)) - \mu_{k-1}(f_{c_{k-1}}(x_{k-1}) - f_{c_{k-1}}(x^*)) + m_k(f_{c_k}(x_k) - f_{c_k}(x^*)) \\ + \nu_k \|x_k - x^*\|^2 - \nu_{k-1} \|x_{k-1} - x^*\|^2 + n_k \|x_k - x^*\|^2 + \eta_k \|x_k - x_{k-1}\|^2 + b_{k-1}^2 \|\nabla f(x_k)\|^2 \leq 0. \end{aligned} \quad (95)$$

Note that $\nu_k \sim C$. Therefore, from (95), similarly as in the proof of Theorem 12, we deduce that $\|x_k - x^*\|$ is bounded, and therefore (x_k) is bounded. Further,

$$\lim_{k \rightarrow +\infty} (\|a(x_k - x^*) + k(x_{k+1} - x_k)\|^2 + \mu_k(f_{c_k}(x_k) - f_{c_k}(x^*)) + \nu_k \|x_k - x^*\|^2) = 0,$$

that is, $\lim_{k \rightarrow +\infty} \nu_k \|x_k - x^*\|^2 = 0$ and hence $\lim_{k \rightarrow +\infty} x_k = x^*$.

Case II. Assume that there exists $k_0 \in \mathbb{N}$ such that $\|x_k\| < \|x^*\|$ for all $k \geq k_0$. From there we get that (x_k) is bounded. Now, take $\bar{x} \in \mathcal{H}$ a weak sequential cluster point of (x_k) , which exists since (x_k) is bounded. This means that there exists a sequence $(k_n)_{n \in \mathbb{N}} \subseteq [k_0, +\infty) \cap \mathbb{N}$ such that $k_n \rightarrow +\infty$ and x_{k_n} converges weakly to \bar{x} as $n \rightarrow +\infty$. Since f is weakly lower semicontinuous, according to Theorem 12 we have $f(\bar{x}) \leq \liminf_{n \rightarrow +\infty} f(x_{k_n}) = \min f$, hence $\bar{x} \in \operatorname{argmin} f$. Since the norm is weakly lower semicontinuous, we deduce that

$$\|\bar{x}\| \leq \liminf_{n \rightarrow +\infty} \|x_{k_n}\| \leq \|x^*\|.$$

According to the definition of x^* , we get $\bar{x} = x^*$. Therefore (x_k) converges weakly to x^* . So

$$\|x^*\| \leq \liminf_{k \rightarrow +\infty} \|x_k\| \leq \limsup_{t \rightarrow +\infty} \|x_k\| \leq \|x^*\|.$$

Therefore, we have $\lim_{k \rightarrow +\infty} \|x_k\| = \|x^*\|$. From the previous relation and the fact that $x_k \rightharpoonup x^*$ as $k \rightarrow +\infty$, we obtain the strong convergence, that is $\lim_{k \rightarrow +\infty} x_k = x^*$.

Case III. Suppose that for every $k \geq k_0$ there exists $l \geq k$ such that $\|x^*\| > \|x_l\|$, and suppose also there exists $m \geq k$ such that $\|x^*\| \leq \|x_m\|$. So, let $k_1 \geq k_0$ and $l_1 \geq k_1$ such that $\|x^*\| > \|x_{l_1}\|$. Let $k_2 > l_1$ and $l_2 \geq k_2$ such that $\|x^*\| > \|x_{l_2}\|$. Continuing the process, we obtain (x_{l_n}) , a subsequence of (x_k) with the property that $\|x_{l_n}\| < \|x^*\|$ for all $n \in \mathbb{N}$. By reasoning as in **Case II**, we obtain that $\lim_{n \rightarrow +\infty} x_{l_n} = x^*$. Consequently, $\liminf_{k \rightarrow +\infty} \|x_k - x^*\| = 0$.

5.3 Non-smooth case

Let us extend the results of the previous sections to the case of a proper lower semicontinuous and convex function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. We rely on the basic properties of the Moreau envelope $f_\lambda : \mathcal{H} \rightarrow \mathbb{R}$ (λ is a positive real parameter), which is defined by

$$f_\lambda(x) = \min_{z \in \mathcal{H}} \left\{ f(z) + \frac{1}{2\lambda} \|z - x\|^2 \right\}, \quad \text{for any } x \in \mathcal{H}.$$

Recall that f_λ is a convex differentiable function, whose gradient is λ^{-1} -Lipschitz continuous, and such that $\min_{\mathcal{H}} f = \min_{\mathcal{H}} f_\lambda$, $\operatorname{argmin}_{\mathcal{H}} f_\lambda = \operatorname{argmin}_{\mathcal{H}} f$. The interested reader may refer to [21, 24] for a comprehensive treatment of the Moreau envelope in a Hilbert setting. Since the set of minimizers is preserved by taking the Moreau envelope, the idea is to replace f by f_λ in the previous algorithm, and take advantage of the fact that f_λ is continuously differentiable. Then, algorithm (IPATRE) applied to f_λ now reads (recall that $\alpha_k = 1 - \frac{\alpha}{k}$)

$$\text{(IPATRE)} \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = \operatorname{prox}_{f_\lambda} \left(y_k - \frac{c}{k^2} x_k \right). \end{cases}$$

By applying Theorems 12 and 13, we obtain fast convergence of the sequence (x_k) to the element of minimum norm of f . Thus, we just need to formulate these results in terms of f and its proximal mapping. This is straightforward thanks to the following formulae from proximal calculus [21]:

1. $f_\lambda(x) = f(\operatorname{prox}_{\lambda f}(x)) + \frac{1}{2\lambda} \|x - \operatorname{prox}_{\lambda f}(x)\|^2$.
2. $\nabla f_\lambda(x) = \frac{1}{\lambda} (x - \operatorname{prox}_{\lambda f}(x))$.
3. $\operatorname{prox}_{\theta f_\lambda}(x) = \frac{\lambda}{\lambda + \theta} x + \frac{\theta}{\lambda + \theta} \operatorname{prox}_{(\lambda + \theta)f}(x)$.

We obtain the following relaxed inertial proximal algorithm (NS stands for non-smooth):

$$\text{(IPATRE-NS)} \quad \begin{cases} y_k = x_k + (1 - \frac{\alpha}{k})(x_k - x_{k-1}) \\ x_{k+1} = \frac{\lambda}{1 + \lambda} \left(y_k - \frac{c}{k^2} x_k \right) + \frac{1}{1 + \lambda} \operatorname{prox}_{(\lambda + 1)f} \left(y_k - \frac{c}{k^2} x_k \right). \end{cases}$$

Theorem 14 *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, proper function. Assume that $\alpha > 3$. Let (x_k) be a sequence generated by (IPATRE-NS). Then for all $s \in [\frac{1}{2}, 1[$, we have:*

- (i) $f(\operatorname{prox}_{\lambda f}(x_k)) - \min_{\mathcal{H}} f = o(k^{-2s})$, $\|x_k - x_{k-1}\| = o(k^{-s})$,
 $\|x_k - \operatorname{prox}_{\lambda f}(x_k)\| = o(k^{-s})$ as $k \rightarrow +\infty$.

$$(ii) \quad \sum_{k=1}^{+\infty} k^{2s-1} (f(\operatorname{prox}_{\lambda f}(x_k)) - \min_{\mathcal{H}} f) < +\infty, \quad \sum_{k=1}^{+\infty} k^{2s-1} \|x_k - x_{k-1}\|^2 < +\infty,$$

$$\sum_{k=1}^{+\infty} k^{2s} \|x_k - \operatorname{prox}_{\lambda f}(x_k)\|^2 < +\infty.$$

(iii) $\liminf_{k \rightarrow +\infty} \|x_k - x^*\| = 0$. Further, (x_k) converges strongly to x^* the element of minimum norm of $\operatorname{argmin} f$, if (x_k) is in the interior of the ball $B(0, \|x^*\|)$ for k large enough, or if (x_k) is in the complement of the ball $B(0, \|x^*\|)$ for k large enough.

6 Conclusion, perspective

In the framework of convex optimization in general Hilbert spaces, we have introduced an inertial dynamic in which the damping coefficient and the Tikhonov regularization coefficient vanish as time tends to infinity. The judicious adjustment of these parameters makes it possible to obtain

trajectories converging quickly (and strongly) towards the minimum norm solution. This seems to be the first time that these two properties have been obtained for the same dynamic. Indeed, the Nesterov accelerated gradient method and the hierarchical minimization attached to the Tikhonov regularization are fully effective within this dynamic. On the basis of Lyapunov's analysis, we have developed an in-depth mathematical study of the dynamic which is a valuable tool for the development of corresponding results for algorithms obtained by temporal discretization. We thus obtained similar results for the corresponding proximal algorithms. This study opens up a large field of promising research concerning first-order optimization algorithms. Many interesting questions such as the introduction of Hessian-driven damping to attenuate oscillations [9], [19], [23], and the study of the impact of errors, perturbations, deserve further study. These results also adapt well to the numerical analysis of inverse problems for which strong convergence and obtaining a solution close to a desired state are key properties.

References

1. F. ALVAREZ, H. ATTOUCH, *Convergence and asymptotic stabilization for some damped hyperbolic equations with non-isolated equilibria*, ESAIM Control Optim. Calc. Var. 6 (2001), 539–552.
2. F. ALVAREZ, A. CABOT, *Asymptotic selection of viscosity equilibria of semilinear evolution equations by the introduction of a slowly vanishing term*, Discrete Contin. Dyn. Syst. 15 (2006), 921–938.
3. V. APIDOPOULOS, J.-F. AUJOL, CH. DOSSAL, *The differential inclusion modeling the FISTA algorithm and optimality of convergence rate in the case $b \leq 3$* , SIAM J. Optim., 28(1) (2018), 551–574.
4. H. ATTOUCH, *Viscosity solutions of minimization problems*, SIAM J. Optim. 6 (3) (1996), 769–806.
5. H. ATTOUCH, R.I. BOŢ, E.R. CSETNEK, *Fast optimization via inertial dynamics with closed-loop damping*, Journal of the European Mathematical Society (JEMS), 2021, hal-02910307.
6. H. ATTOUCH, L.M. BRICEÑO-ARIAS, P.L. COMBETTES, *A parallel splitting method for coupled monotone inclusions*, SIAM J. Control Optim. 48 (5) (2010), 3246–3270.
7. H. ATTOUCH, L.M. BRICEÑO-ARIAS, P.L. COMBETTES, *A strongly convergent primal-dual method for nonoverlapping domain decomposition*, Numerische Mathematik, 133(3) (2016), 443–470.
8. H. ATTOUCH, A. CABOT, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, J. Differential Equations, 263 (9), (2017), 5412–5458.
9. H. ATTOUCH, Z. CHBANI, J. FADILI, H. RIAHI, *First order optimization algorithms via inertial systems with Hessian driven damping*, Math. Program. (2020), <https://doi.org/10.1007/s10107-020-01591-1>.
10. H. ATTOUCH, Z. CHBANI, J. PEYPOUQUET, P. REDONT, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Math. Program., 168 (1-2) (2018), 123–175.
11. H. ATTOUCH, Z. CHBANI, H. RIAHI, *Combining fast inertial dynamics for convex optimization with Tikhonov regularization*, J. Math. Anal. Appl, 457 (2018), 1065–1094.
12. H. ATTOUCH, R. COMINETTI, *A dynamical approach to convex minimization coupling approximation with the steepest descent method*, J. Differential Equations, 128 (2) (1996), 519–540.
13. H. ATTOUCH, M.-O. CZARNECKI, *Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria*, J. Differential Equations 179 (2002), 278–310.
14. H. ATTOUCH, M.-O. CZARNECKI, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, J. Differential Equations 248 (2010), 1315–1344.
15. H. ATTOUCH, M.-O. CZARNECKI, J. PEYPOUQUET, *Prox-penalization and splitting methods for constrained variational problems*, SIAM J. Optim. 21 (2011), 149–173.
16. H. ATTOUCH, M.-O. CZARNECKI, J. PEYPOUQUET, *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM J. Optim. 21 (2011), 1251–1274.
17. H. ATTOUCH, M.-O. CZARNECKI, *Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects*, J. Differential Equations, 262 (3) (2017), 2745–2770.
18. H. ATTOUCH, J. PEYPOUQUET, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$* , SIAM J. Optim., 26(3) (2016), pp. 1824–1834.
19. H. ATTOUCH, J. PEYPOUQUET, P. REDONT, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Differential Equations, 261(10), (2016), 5734–5783.
20. J.-B. BAILLON, R. COMINETTI, *A convergence result for non-autonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming*, J. Funct. Anal. 187 (2001) 263–273.
21. H. BAUSCHKE, P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).
22. R. I. BOŢ, E. R. CSETNEK, *Forward-Backward and Tseng's type penalty schemes for monotone inclusion problems*, Set-Valued Var. Anal. 22 (2014), 313–331.

23. R. I. BOŢ, E. R. CSETNEK, S.C. LÁSZLÓ, *Tikhonov regularization of a second order dynamical system with Hessian damping*, Math. Program. (2020), <https://doi.org/10.1007/s10107-020-01528-8>.
24. H. BRÉZIS, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, (1972).
25. A. CABOT, *Inertial gradient-like dynamical system controlled by a stabilizing term*, J. Optim. Theory Appl. 120 (2004) 275–303.
26. A. CABOT, *Proximal point algorithm controlled by a slowly vanishing term: Applications to hierarchical minimization*, SIAM J. Optim. 15 (2) (2005), 555–572.
27. A. CABOT, H. ENGLER, S. GADAT, *On the long time behavior of second order differential equations with asymptotically small dissipation* Trans. Amer. Math. Soc. 361 (2009), 5983–6017.
28. A. CHAMBOLLE, CH. DOSSAL, *On the convergence of the iterates of Fista*, J. Opt. Theory Appl., 166 (2015), 968–982.
29. R. COMINETTI, *Coupling the proximal point algorithm with approximation methods*, J. Optim. Theory Appl. 95 (3) (1997), 581–600.
30. R. COMINETTI, J. PEYPOUQUET, S. SORIN, *Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization*, J. Differential Equations, 245 (2008), 3753–3763.
31. A. FIACCO, G. MCCORMICK, *Nonlinear programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York, (1968).
32. A. HARAUX, M.A. JENDOUBI, *A Liapunov function approach to the stabilization of second-order coupled systems*, (2016) arXiv preprint arXiv:1604.06547.
33. S.A. HIRSTOAGA, *Approximation et résolution de problèmes d'équilibre, de point fixe et d'inclusion monotone*. PhD thesis, Université Pierre et Marie Curie - Paris VI, 2006, HAL Id: tel-00137228.
34. M.A. JENDOUBI, R. MAY, *On an asymptotically autonomous system with Tikhonov type regularizing term*, Archiv der Mathematik 95 (4) (2010), 389–399.
35. Y. NESTEROV, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$* , Soviet Math. Dokl. 27 (1983), 372–376.
36. Y. NESTEROV, *Introductory lectures on convex optimization: A basic course*, volume 87 of Applied Optimization. Kluwer Academic Publishers, Boston, MA, 2004.
37. B. POLYAK, *Introduction to Optimization*, New York, NY: Optimization Software-Inc, 1987.
38. W. SU, S. BOYD, E. J. CANDÈS, *A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights*. NIPS, December 2014.
39. A. N. TIKHONOV, Doklady Akademii Nauk SSSR 151 (1963) 501–504, (Translated in "Solution of incorrectly formulated problems and the regularization method". Soviet Mathematics 4 (1963) 1035–1038).
40. A. N. TIKHONOV, V. Y. ARSENIN, *Solutions of Ill-Posed Problems*, Winston, New York, 1977.